

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200580044434.0

[51] Int. Cl.

C12N 9/22 (2006.01)

C12P 21/02 (2006.01)

[43] 公开日 2007 年 12 月 12 日

[11] 公开号 CN 101087875A

[22] 申请日 2005.12.1

[21] 申请号 200580044434.0

[30] 优先权

[32] 2004.12.22 [33] US [31] 60/639,146

[86] 国际申请 PCT/US2005/043603 2005.12.1

[87] 国际公布 WO2006/068802 英 2006.6.29

[85] 进入国家阶段日期 2007.6.22

[71] 申请人 AMBRX 公司

地址 美国加利福尼亚州

[72] 发明人 安德鲁·鲍赛尔 丘霍松

[74] 专利代理机构 北京科龙寰宇知识产权代理有限
责任公司

代理人 孙皓晨

权利要求书 3 页 说明书 69 页 序列表 4 页
附图 3 页

[54] 发明名称

氨酰基 - tRNA 合成酶的组合物及其用途

[57] 摘要

本发明提供生产包括正交 tRNA、正交氨酰基 - tRNA 合成酶和 tRNA/合成酶的正交对的蛋白质生物合成器的组分的组合物和方法。本发明也提供用于鉴别所述正交对的方法以及使用所述正交对来生产蛋白质的方法。

1. 一种组合物，其包含具有选自由以下序列组成的群组的核酸序列的氨酰基 tRNA 合成酶 (RS)：SEQ ID NO : 4、编码 SEQ ID NO : 5 的氨基酸序列的多核苷酸序列、编码 SEQ ID NO : 17 的氨基酸序列的多核苷酸序列、其互补多核苷酸序列和其保守变化。
2. 根据权利要求 1 所述的组合物，其中所述 RS 优选氨基酰化 tRNA。
3. 根据权利要求 2 所述的组合物，其中所述 tRNA 是经非天然编码的氨基酸氨基酰化。
4. 根据权利要求 3 所述的组合物，其中所述非天然编码的氨基酸是对乙酰基苯丙氨酸。
5. 一种组合物，其包含根据权利要求 1 所述的 RS 和 tRNA，其中所述 tRNA 是经化学氨基酰化的。
6. 一种组合物，其包含根据权利要求 1 所述的 RS 和 tRNA，其中所述 tRNA 是经酶促氨基酰化的。
7. 根据权利要求 6 所述的组合物，其中所述 tRNA 是通过所述 RS 酶促氨基酰化。
8. 根据权利要求 6 所述的组合物，其中所述 tRNA 是通过核糖酶酶促氨基酰化。
9. 根据权利要求 1 所述的组合物，其另外包含 tRNA，其中所述 tRNA 是通过所述 RS 用氨基酸氨基酰化。
10. 根据权利要求 9 所述的组合物，其中所述氨基酸是非天然编码的氨基酸。
11. 根据权利要求 10 所述的组合物，其中所述非天然编码的氨基酸是对乙酰基苯丙氨酸。
12. 根据权利要求 9 所述的组合物，其中所述 tRNA 是由原始 tRNA 得到。
13. 根据权利要求 9 所述的组合物，其中所述 tRNA 是由詹氏甲烷球菌 (*M. janneschii*) 得到。
14. 根据权利要求 1 所述的组合物，其另外包含转译系统。
15. 根据权利要求 14 所述的组合物，其中所述转译系统是无细胞转译系统。
16. 根据权利要求 14 所述的组合物，其中所述转译系统是细胞溶胞产物。
17. 根据权利要求 14 所述的组合物，其中所述转译系统是重构系统。
18. 根据权利要求 14 所述的组合物，其中所述转译系统是细胞转译系统。
19. 一种细胞，其包含转译系统，其中所述转译系统包含根据权利要求 1 所述的 RS。
20. 根据权利要求 19 所述的细胞，其中所述细胞是真核细胞。

21. 根据权利要求 20 所述的细胞，其中所述真核细胞是酵母细胞。
22. 根据权利要求 20 所述的细胞，其中所述真核细胞是真菌细胞。
23. 根据权利要求 20 所述的细胞，其中所述真核细胞是哺乳动物细胞。
24. 根据权利要求 20 所述的细胞，其中所述真核细胞是昆虫细胞。
25. 根据权利要求 20 所述的细胞，其中所述真核细胞是植物细胞。
26. 根据权利要求 19 所述的细胞，其中所述细胞是非真核细胞。
27. 根据权利要求 26 所述的细胞，其中所述非真核细胞是大肠杆菌 (*E. coli*) 细胞。
28. 根据权利要求 19 所述的细胞，其另外包含通过所述 RS 氨基酰化的 tRNA 和编码所关注的多肽的多核苷酸，其中所述多核苷酸包含通过所述 tRNA 识别的选择密码子。
29. 根据权利要求 28 所述的细胞，其中所述所关注的多肽是人类生长激素。
30. 一种大肠杆菌细胞，其包含根据权利要求 1 所述的 RS。
31. 一种酵母细胞，其包含根据权利要求 1 所述的 RS。
32. 一种真菌细胞，其包含根据权利要求 1 所述的 RS。
33. 一种哺乳动物细胞，其包含根据权利要求 1 所述的 RS。
34. 一种昆虫细胞，其包含根据权利要求 1 所述的 RS。
35. 一种植物细胞，其包含根据权利要求 1 所述的 RS。
36. 一种载体，其包含编码具有选自由以下序列组成的群组的核酸序列的 RS 的多核苷酸：SEQ ID NO : 4、编码 SEQ ID NO : 5 的氨基酸序列的多核苷酸序列、编码 SEQ ID NO : 17 的氨基酸序列的多核苷酸序列、其互补多核苷酸序列和其保守变化。
37. 根据权利要求 36 所述的载体，其中所述载体包含质粒、粘质粒、噬菌体或病毒。
38. 根据权利要求 36 所述的载体，其中所述载体是表达载体。
39. 一种细胞，其包含根据权利要求 36 所述的载体。
40. 一种在细胞中用所选氨基酸在规定位置上生产多肽的方法，所述方法包含：
使所述细胞在适当培养基中生长，其中所述细胞包含包括至少一个选择密码子并编码多肽的核酸；和，
提供所述所选氨基酸；
其中所述细胞另外包含：
在所述细胞中起作用的根据权利要求 1 所述的正交 RS (O-RS)，其中所述 RS 具有选自由以下序列组成的群组的氨基酸序列：SEQ ID NO : 5、SEQ ID NO : 17、其互补多核苷酸序列和其保守变化；和，
识别出所述选择密码子的正交 tRNA (O-tRNA)，其中所述 O-RS 用所述所选氨

基酸氨基酰化所述 O-tRNA。

41. 根据权利要求 40 所述的方法，其中所述所选氨基酸是对乙酰基苯丙氨酸。

42. 一种通过根据权利要求 40 所述的方法产生的多肽。

43. 根据权利要求 40 所述的多肽，其中所述所选氨基酸是对乙酰基苯丙氨酸。

44. 根据权利要求 40 所述的方法，其中所述多肽是人类生长激素。

45. 一种多核苷酸，其具有选自由以下序列组成的群组的核酸序列：SEQ ID No. 4、编码 SEQ ID NO: 5 的氨基酸序列的多核苷酸序列、编码 SEQ ID NO: 17 的氨基酸序列的多核苷酸序列、其互补多核苷酸序列和其保守变化。

氨酰基-tRNA 合成酶的组合物及其用途

交叉引用的相关申请案

本申请案要求 2004 年 12 月 22 日提交的美国临时专利申请案第 60/639,146 号的优先权，其说明书全部并入本文中。

技术领域

本发明属于转译生物化学领域。本发明涉及用于生产氨酰基-tRNA 合成酶的组合物的方法和氨酰基-tRNA 合成酶的组合物及其用途。本发明也涉及使用所述氨酰基-tRNA 合成酶和相关组合物在细胞中生产蛋白质的方法。

背景技术

从细菌到人类的每种已知生物体的遗传密码编码相同的 20 种常见氨基酸。相同的 20 种天然氨基酸的不同组合形成蛋白质，所述蛋白质实际上执行从光合作用到信号转导和免疫反应的生命的的所有复杂过程。为了研究和修饰蛋白质结构和功能，科学家已尝试操纵蛋白质的遗传密码和氨基酸序列。然而，仍难以除去由遗传密码施加的约束，其使蛋白质受限于 20 种遗传编码的标准构造单元 (building block) (硒代半胱氨酸 (参见 (例如) A. Bock 等人, (1991), Molecular Microbiology 5:515-20) 和吡咯赖氨酸 (参见 (例如) G. Srinivasan 等人, (2002), Science 296:1459-62) 为少有的例外)。

对于除去所述约束已经取得了一些进展，尽管所述进展是有限的并且合理地控制蛋白质结构和功能的能力仍处于其初期。举例而言，化学工作者已开发出合成和操纵小分子结构的方法和策略 (参见 (例如) E. J. Corey 和 X.-M. Cheng, The Logic of Chemical Synthesis (Wiley-Interscience, New York, 1995))。全合成 (参见 (例如) B. Merrifield, (1986), Science 232:341-7 (1986)) 和半合成方法 (参见 (例如) D. Y. Jackson 等人, (1994) Science 266:243-7; 和 P. E. Dawson 和 S. B. Kent, (2000), Annual Review of Biochemistry 69:923-60), 已使合成肽和小蛋白质成为可能，但是所述方法对超过 10 千道尔顿 (kDa) 的蛋白质来说，具有有限的实用性。虽然突变方法是有效的，但是其受限于有限数量的结构变化。在许多情况下，已经有可能在整个蛋白质中竞争性地并入常见氨基酸的封闭结构类似物。参见 (例如), R. Furter, (1998), Protein Science 7:419-26; K. Kirshenbaum 等人, (2002), ChemBioChem 3:235-7; 和 V. Doring 等人, (2001), Science 292:501-4。化

学肽连接和天然化学连接描述于美国专利第 6,184,344 号、美国专利公开案第 2004/0138412 号、美国专利公开案第 2003/0208046 号、WO 02/098902 和 WO 03/042235 中,所述专利是以引用的方式并入本文中。Lu 等人在 *Mol Cell*. 2001 年 10 月; 8 (4) :759-69 中描述了一种方法,其中蛋白质与含有非天然氨基酸的合成肽化学连接(经表达蛋白质连接)。

早期的工作证明大肠杆菌 (*E. coli*) 的转译机器将容纳结构上与常见 20 种氨基酸相似的氨基酸。参见, Hortin, G.和 Boime, I. (1983) Methods Enzymol. 96:777-784。所述工作另外通过放宽内源性大肠杆菌合成酶的特异性,以便其活化非天然氨基酸以及其同源天然氨基酸而扩展。此外,展示了编辑域中的突变也可用以扩展内源性合成酶的底物范畴。参见, Doring, V.等人, (2001) Science 292:501-504。然而,所述策略都受限于重新编码遗传密码而不是扩展遗传密码,并且导致常见 20 种氨基酸中之一者经非天然氨基酸不同程度地取代。

新近,展示了非天然氨基酸可通过将经化学氨基酰化的正交琥珀抑制基因 tRNA 添加到活体外转录/转译反应中,而得以在活体外部位特异性地并入蛋白质中。参见(例如), Noren, C. J.等人, (1989) Science 244:182-188; Bain, J. D.等人, (1989) J. Am. Chem. Soc. 111:8013-8014; Dougherty, D. A. (2000) Curr. Opin. Chem. Biol. 4, 645-652; Cornish, V. W.等人, (1995) Angew. Chem., Int. Ed. 34:621-633; J. A. Ellman 等人, (1992), Science 255:197-200; 和 D. Mendel 等人, (1995), Annual Review of Biophysics and Biomolecular Structure 24:435-462。所述研究展示,核糖体和转译因子可与大量非天然氨基酸,甚至是具有异常结构的所述非天然氨基酸相容。不幸地,tRNA 的化学氨基酰化是困难的,并且所述工艺的化学计量性质严重地限制可产生的蛋白质的量。

已经将非天然氨基酸微量注射到细胞中。举例而言,非天然氨基酸通过微量注射经化学错误酰化的嗜热四膜虫 (*Tetrahymena thermophila*) tRNA (例如, M.E. Saks 等人, (1996), *An engineered Tetrahymena tRNA^{Gln} for in vivo incorporation of unnatural amino acids into proteins by nonsense suppression*, J. Biol. Chem. 271:23169-23175) 和相关 mRNA 而引入爪蟾卵母细胞 (*Xenopus oocyte*) 中的烟碱酸乙酰胆碱受体中(例如, M.W. Nowak 等人, (1998), *In vivo incorporation of unnatural amino acids into ion channels in Xenopus oocyte expression system*, Method Enzymol. 293:504-529)。又参见, D. A. Dougherty (2000), *Unnatural amino acids as probes of protein structure and function*, Curr. Opin. Chem. Biol. 4:645-652 和 M. W. Nowak, P. C. Kearney, J. R. Sampson, M. E. Saks, C. G. Labarca, S. K. Silverman, W. G. Zhong, J. Thorson, J. N. Abelson, N. Davidson, P. G.

Schultz, D. A. Dougherty 和 H. A. Lester, Science, 268:439 (1995)。爪蟾卵母细胞与以下活体外制得的两种 RNA 物质共同注射：编码在所关注的氨基酸位置处具有 UAG 终止密码子的目标蛋白质的 mRNA，和经所要的非天然氨基酸氨基酰化的琥珀抑制基因 tRNA。然后，卵母细胞的转译机器在规定的位置上通过 UAG 将非天然氨基酸插入。不幸地，所述方法受限于在细胞中可微量注射的蛋白质，并且因为相关的 tRNA 经活体外化学酰化并且不能再酰化，所以蛋白质的产量极低。

为了克服所述限制，将例如正交 tRNA、正交氨酰基-tRNA 合成酶及其对的新颖组分添加到原核生物大肠杆菌 (*Escherichia coli*, *E. coli*) (参见(例如)L. Wang 等人, (2001), Science 292:498-500) 和真核生物酿酒酵母 (*Saccharomyces cerevisiae*, *S. cerevisiae*) (例如, J. Chin 等人, Science 301:964-7 (2003)) 的蛋白质生物合成机器中，其已使非遗传编码的氨基酸能活体内并入蛋白质中。使用所述方法，响应(例如)琥珀密码子 (TAG) 已经将具有新颖化学、物理或生物性质，包括光亲和标记和可光致异构氨基酸、光致交联氨基酸 (参见(例如), Chin, J. W.等人(2002) Proc. Natl. Acad. Sci. U. S. A. 99:11020-11024; 和 Chin, J. W.等人(2002) J. Am. Chem. Soc. 124:9026-9027)、酮氨基酸 (参见(例如), Wang, L.等人, (2003) Proc. Natl. Acad. Sci. U. S. A. 100:56-61 和 Zhang, Z.等人, Biochem. 42 (22) :6735-6746 (2003))、含有重原子的氨基酸和经糖基化的氨基酸的大量新颖氨基酸有效地且以高保真度并入大肠杆菌和酵母中的蛋白质中。参见(例如), J. W. Chin 和 P. G. Schultz, (2002), ChemBioChem 3(11): 1135-1137; 和 L. Wang 和 P. G. Schultz, (2002), Chem. Comm. 1:1-11。

已经报道了若干其他的正交对。从酿酒酵母 tRNA 和合成酶得到的谷氨酰胺酰基(参见(例如), Liu, D. R.和 Schultz, P. G. (1999) Proc. Natl. Acad. Sci. U. S. A. 96:4780-4785)、天冬氨酰基(参见(例如), Pastrnak, M.等人, (2000) Helv. Chim. Acta 83:2277-2286) 和酪氨酰基(参见(例如), Ohno, S.等人, (1998) J. Biochem. (Tokyo, Jpn.) 124:1065-1068; 和 Kowal, A. K.等人, (2001) Proc. Natl. Acad. Sci. U. S. A. 98:2268-2273) 系统已经被描述用于将非天然氨基酸潜在地并入大肠杆菌中。从大肠杆菌谷氨酰胺酰基(参见(例如), Kowal, A. K.等人, (2001) Proc. Natl. Acad. Sci. U. S. A. 98:2268-2273) 和酪氨酰基(参见(例如), Edwards, H.和 Schimmel, P. (1990) Mol. Cell. Biol. 10:1633-1641) 合成酶得到的系统已经被描述适用于酿酒酵母。大肠杆菌酪氨酰基系统已经用于将 3-碘-L-酪氨酸活体内并入哺乳动物细胞中。参见, Sakamoto, K.等人, (2002) Nucleic Acids Res. 30:4692-4699。通常，所述系统利用琥珀终止密码子。为进一步扩展遗传密码，需要开发生物合成机器的改善和/或其他组分，例如氨酰基-tRNA 合成酶。如基于以下揭示案的

评述将明显可见，本发明满足所述的和其他的需要。

发明内容

为了扩展遗传密码，本发明提供正交氨酰基-tRNA 合成酶的组合物和生产正交氨酰基-tRNA 合成酶的方法。本发明的氨酰基-tRNA 合成酶用非天然编码的氨基酸氨酰化 tRNA。这些转译组分可用来响应由 tRNA 识别出的选择密码子（selector codon）而将所选氨基酸并入生长多肽链中的特定位置中（在核酸转译期间）。

在细胞中用所选氨基酸在规定位置生产蛋白质的方法也是本发明的特征。举例而言，方法包括在适当培养基中使细胞生长，其中所述细胞包含至少一个选择密码子并且编码蛋白质的核酸；和提供所选氨基酸。细胞另外包含：在细胞中起作用并且识别选择密码子的正交 tRNA（O-tRNA）；和用所选氨基酸优先氨酰化 O-tRNA 的正交氨酰基-tRNA 合成酶（O-RS）。通常，O-tRNA 包含在同源合成酶的存在下抑制活性。由所述方法生产的蛋白质也是本发明的特征。

附图说明

图 1-展示具有 TYC 茎突变部位的 J17 tRNA 的四叶式交叉（cloverleaf）结构。

图 2-展示使用 J17 或 J17 突变体（F12、F13、F14）和 E9 RS 的人类生长激素的琥珀突变的抑制。各个样品总细胞溶胞产物通过 SDS PAGE 来分析。

图 3-展示在不同细胞系中，使用 F13 和 E9 RS 的人类生长激素的琥珀突变的抑制。

具体实施方式

定义

在详细地描述本发明之前，应了解，本发明不受限于具体的生物系统，本发明当然可以变化。也应了解，本文中使用的术语仅是为了描述具体的实施例，并且不希望限制本发明的范畴，本发明的范畴将仅由附加权利要求书来限制。如本文中和附加权利要求书中所使用，除非上下文另外清楚地规定，否则单数形式“一”和“所述”包括复数个参考物。因此，举例而言，提及“一种细胞”包括两种或两种以上细胞的组合并且包括所属领域的技术人员已知的其等价物，等等。提及“细菌”包括细菌的混合物诸如此类。

除非在本文中和以下说明书的剩余部分中规定，否则本文中使用的所有技术和科学术语具有与本发明所属领域的技术人员通常所了解的相同的含义。

本文中提及的所有公开案和专利是以引用的方式并入本文中，以达到描述和揭示（例如）公开案中描述的，可能与当前描述的发明联系使用的构筑体和方法的目的。所提供的本文中讨论的公开案仅为在本专利申请案的申请日之前的其揭示案。在此并非解

释为允许本发明者由于先前发明或由于任何其他原因而比所述揭示案提前得到授权。

同源 (Homologous): 蛋白质和/或蛋白质序列, 当其天然地或人工地从共同的原始蛋白质或蛋白质序列得到时, 是“同源的”。同样地, 核酸和/或核酸序列, 当其天然地或人工地从共同的原始核酸或核酸序列得到时, 是同源的。举例而言, 任何天然存在的核酸可通过任何可用的突变方法修饰以包括一个或一个以上选择密码子。当被表达时, 所述经突变的核酸编码包含一种或一种以上所选氨基酸 (例如非天然氨基酸) 的多肽。当然, 突变处理可另外改变一个或一个以上标准密码子, 进而也改变所得突变蛋白质中的一种或一种以上标准氨基酸。所述一种或一种以上标准氨基酸可以变成非天然氨基酸或天然氨基酸。同源性通常从两种或两种以上的核酸或蛋白质 (或其序列) 之间的序列相似性来推断。适用于建立同源性的序列之间相似性的精确百分比随所讨论的核酸和蛋白质而变化, 但仅仅 25% 的序列相似性通常用以建立同源性。例如 30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98% 或 99% 或 99% 以上的高水平的序列相似性也可用以建立同源性。用于确定序列相似性百分比的方法 (例如, 使用缺省参数的 BLASTP 和 BLASTN) 描述于本文中并且通常可用。

正交: 如本文中所使用, 术语“正交”指的是通过所关注的系统 (例如转译系统, 例如细胞) 以降低的效率使用的分子 (例如正交 tRNA (O-tRNA) 和/或正交氨酰基 tRNA 合成酶 (O-RS))。正交指的是正交 tRNA 和/或正交 RS 在所关注的转译系统中不能起作用或效率降低, 例如小于 20% 有效、小于 10% 有效、小于 5% 有效或例如小于 1% 有效。举例而言, 当与内源性 tRNA 通过内源性 RS 的氨酰化相比较时, 所关注的转译系统中的正交 tRNA 是通过所关注的转译系统的任何内源性 RS 以降低或甚至为零的效率氨酰化。在另一实例中, 当与内源性 tRNA 通过内源性 RS 的氨酰化相比较时, 正交 RS 以降低或甚至为零的效率氨酰化所关注的转译系统中的任何内源性 tRNA。可将与第一正交分子一同起作用的第二正交分子引入细胞中。举例而言, 正交 tRNA/RS 对包括所引入的互补组分, 所述互补组分在细胞中以某一效率 (例如, 约 50% 效率、约 60% 效率、约 70% 效率、约 75% 效率、约 80% 效率、约 85% 效率、约 90% 效率、约 95% 效率或约 99% 或 99% 以上的效率) 一同对对应的 tRNA/RS 内源性对的组分起作用。“正交性的改善”指的是与起始物质或天然存在的 tRNA 或 RS 相比较正交性增强。

同源物: 术语“同源物”指的是一同起作用的组分, 例如 tRNA 和氨酰基-tRNA 合成酶。所述组分也可称为互补组分。

优先氨酰化: 术语“优先氨酰化”指的是与 O-RS 氨酰化天然存在的 tRNA

或用以产生 O-tRNA 的起始物质的效率相比较，O-RS 用例如非天然氨基酸的所选氨基酸氨基酰化 O-tRNA 的效率，例如约 70%有效、约 75%有效、约 80%有效、约 85%有效、约 90%有效、约 95%有效或约 99%或 99%以上有效。因而，非天然氨基酸以高保真度，（例如）以对给定选择密码子来说大于约 70%的效率、对给定选择密码子来说大于约 75%的效率、对给定选择密码子来说大于约 80%的效率、对给定选择密码子来说大于约 85%的效率、对给定选择密码子来说大于约 90%的效率、对给定选择密码子来说大于约 95%的效率或对给定选择密码子来说大于约 99%的效率，并入生长多肽链中。

选择密码子：术语“选择密码子”指的是在转译过程中通过 O-tRNA 识别并且不通过内源性 tRNA 识别的密码子。O-tRNA 反密码子环识别 mRNA 上的选择密码子并且在多肽中的所述部位处并入其氨基酸，例如所选氨基酸，诸如非天然氨基酸。选择密码子可包括（但不限于）（例如）无义密码子，诸如包括（但不限于）琥珀密码子、赭石密码子和蛋白石密码子的终止密码子；4 或 4 个以上碱基密码子；稀有密码子；由天然或非天然碱基对得到的密码子和/或类似物。对给定系统来说，选择密码子也可包括天然 3 碱基密码子之一，其中内源性系统不使用（或很少使用）所述天然 3 碱基密码子。举例而言，内源性系统包括缺乏识别天然 3 碱基密码子的 tRNA 的系统和/或其中天然 3 碱基密码子为稀有密码子的系统。

抑制基因 tRNA：抑制基因 tRNA 是在给定转译系统中，（例如）通过提供响应选择密码子而将氨基酸并入多肽链中的机制，来改变信使 RNA（mRNA）的读取的 tRNA。举例而言，抑制基因 tRNA 可读遍包括（但不限于）终止密码子、4 碱基密码子或稀有密码子的密码子。

抑制活性：术语“抑制活性”指的是例如抑制基因 tRNA 的 tRNA 读遍选择密码子的能力。活性可表示为当与对照物（例如缺乏同源物合成酶）相比时，所观察到的活性的百分比。

转译系统：术语“转译系统”指的是将天然存在的氨基酸并入生长多肽链（蛋白质）中所必需的组分。转译系统的组分可包括（例如）核糖体、tRNA、合成酶、mRNA，诸如此类。可将本发明的组分添加到活体外或活体内转译系统中。转译系统的实例包括（但不限于）非真核细胞，例如细菌（诸如大肠杆菌）；真核细胞，例如酵母细胞、哺乳动物细胞、植物细胞、藻类细胞、真菌细胞、昆虫细胞；无细胞的转译系统，例如细胞溶胞产物；和/或类似物。

转译系统可以是细胞的或无细胞的，并且可以是原核的或真核的。细胞转译系统包括（但不限于）全细胞制剂，诸如透性化细胞或细胞培养物，其中所要的核酸序列可转

录成 mRNA 并且所述 mRNA 被转译。无细胞转译系统为市售的并且许多不同类型和系统是熟知的。无细胞系统的实例包括（但不限于）原核溶胞产物，诸如大肠杆菌溶胞产物；和真核溶胞产物，诸如小麦胚芽提取物、昆虫细胞溶胞产物、兔网织红细胞溶胞产物、兔卵母细胞溶胞产物和人类细胞溶胞产物。当所得蛋白质经糖基化、磷酸化或另外修饰时，真核提取物或溶胞产物可为优选的，因为许多所述修饰仅在真核系统中为可能的。一些所述提取物和溶胞产物为市售的（Promega; Madison, Wis.; Stratagene; La Jolla, Calif; Amersham; Arlington Heights, Ill.; GIBCO/BRL; Grand Island, N.Y.）。诸如含有微粒体膜的犬胰腺提取物的膜状提取物也可用，其适用于转译分泌蛋白。

也可以使用重构转译系统。经纯化的转译因子的混合物以及由诸如起始因子-1 (IF-1)、IF-2、IF-3 (α 或 β)、延伸因子 T (EF-Tu) 或末端因子的经纯化的转译因子补充的溶胞产物的组合或溶胞产物也已经成功用以将 mRNA 转译成蛋白质。无细胞系统也可以是偶联的转录/转译系统，其中将 DNA 引入系统中，转录成 mRNA 并且将 mRNA 如 *Current Protocols in Molecular Biology* (F. M. Ausubel 等编, Wiley Interscience, 1993) 中所述来转译，该文献在此特别以引用的方式并入。在真核转录系统中转录的 RNA 可以呈异核 RNA (hnRNA) 或 5'-末端帽 (7-甲基鸟嘌呤核苷) 和 3'-末端多聚 A 有尾成熟 mRNA 形式，其可为某些转译系统中的优点。举例而言，加帽 mRNA 以高效率在网织红细胞溶胞产物系统中被转译。

所选氨基酸：术语“所选氨基酸”指的是任何所要的天然存在的氨基酸或非天然氨基酸。如本文中所使用，术语“非天然氨基酸”或“非天然编码的氨基酸”指的是任何氨基酸、经修饰的氨基酸和/或不为 20 种常见天然存在的氨基酸之一或硒代半胱氨酸或吡咯赖氨酸的氨基酸类似物。可以与术语“非天然编码的氨基酸”和“非天然氨基酸”同义地使用的其他术语为“非天然的氨基酸”、“非天然存在的氨基酸”和其各种用连字号连接的和未用连字号连接的型式。术语“非天然编码的氨基酸”也包括（但不限于）通过修饰（例如，后转译修饰）天然编码的氨基酸（包括但不限于 20 种常见氨基酸或吡咯赖氨酸和硒代半胱氨酸）产生的氨基酸，但不为通过转译复合物而将其本身天然地并入生长多肽链的氨基酸。所述非天然存在的氨基酸的实例包括（但不限于）N-乙酰氨基葡萄糖基-L-丝氨酸、N-乙酰氨基葡萄糖基-L-苏氨酸和 O-磷酸化酪氨酸。

由...得到：如本文中所使用，术语“由...得到”指的是从规定分子或生物体分离的组分或使用来自规定分子或生物体信息制得的组分。

正性选择或筛选标记物：如本文中所使用，术语“正性选择或筛选标记物”指的是当存在时（例如）经表达、经活化等等时使得具有正性选择标记物的细胞从不具有正性

选择标记物的细胞鉴别出的标记物。

负性选择或筛选标记物：如本文中所使用，术语“负性选择或筛选标记物”指的是当存在时（例如）经表达、经活化等等时容许鉴别出不具有所要性质的细胞（例如，当与具有所要性质的细胞相比时）的标记物。

报告基因 (Reporter)：如本文中所使用，术语“报告基因”指的是可用以选择所关注系统的目标组分的组分。举例而言，报告基因可包括蛋白质，例如给与抗生素抗性或敏感性的酶（包括（但不限于） β -内酰胺酶、氯霉素乙酰转移酶（CAT），诸如此类）、荧光筛选标记物（包括（但不限于）绿色荧光蛋白质（例如 GFP）、YFP、EGFP、RFP）、发光标记物（包括（但不限于）萤火虫荧光素酶蛋白质）、基于亲和力的筛选标记物；或正性或负性可选择标记物基因，诸如 lacZ、 β -gal/lacZ（ β -半乳糖苷酶）、ADH（醇脱氢酶）、his3、ura3、leu2、lys2 等等。

真核生物：如本文中所使用，术语“真核生物”指的是属于真核种系发生域（phylogenetic domain Eucarya）的生物体，诸如动物（包括（但不限于）哺乳动物、昆虫、爬虫、鸟等），纤毛虫，植物（包括（但不限于）单子叶植物、双子叶植物、藻类等），真菌，酵母，鞭毛虫，微粒子虫目，原生生物等。

非真核生物：如本文中所使用，术语“非真核生物”指的是非真核生物体。举例而言，非真核生物体可属于真细菌类（Eubacteria）（包括（但不限于）大肠杆菌、嗜热细菌（*Thermus thermophilus*）、嗜热脂肪芽孢杆菌（*Bacillus stearothermophilus*）、荧光极毛杆菌（*Pseudomonas fluorescens*）、绿脓杆菌（*Pseudomonas aeruginosa*）、恶息极毛杆菌（*Pseudomonas putida*）等）种系发生域，或原生类（Archaea）（例如，詹氏甲烷球菌（*Methanococcus jannaschii*）、横川病毒（*Methanobacterium thermoautotrophicum*）、诸如富饶盐菌（*Haloferax volcanii*）和盐杆菌种（*Halobacterium species*）NRC-1 的盐杆菌属（*Halobacterium*）、闪烁古生球菌（*Archaeoglobus fulgidus*）、火球菌（*Pyrococcus furiosus*）、掘越氏热球菌（*Pyrococcus horikoshii*）、嗜热菌敏捷气热菌（*Aeuropyrum pernix*）等）种系发生域。

保守变体：术语“保守变体”指的是例如保守变体 O-tRNA 或保守变体 O-RS 的转译组分，所述转译组分如例如 O-tRNA 或 O-RS 的组分（保守变体以其为基础）一样执行功能，但在序列中具有变化。举例而言，尽管 O-tRNA 和保守变体 O-tRNA 不具有相同序列，但是 O-RS 将用例如非天然氨基酸的所选氨基酸氨基酰化互补的 O-tRNA 或保守变体 O-tRNA。同样地，尽管 O-RS 和保守变体 O-RS 不具有相同序列，但是 tRNA 将通过互补的 O-RS 或保守变体 O-RS，经例如非天然氨基酸的所选氨基酸氨基酰化。只要

保守变体与对应的 O-tRNA 或 O-RS 互补，保守变体就可在序列中具有（例如）1 种变化、2 种变化、3 种变化、4 种变化或 5 种或 5 种以上的变化。

选择或筛选剂：如本文中所使用，术语“选择或筛选剂”指的是当存在时，允许从群体中选择/筛选某些组分的试剂。举例而言，选择或筛选剂包括（但不限于）（例如）营养素、抗生素、光的波长、抗体、经表达的多核苷酸等等。选择剂可（例如）因浓度、强度等而变化。

术语“未有效地识别出”指的是来自一种生物体的 RS 氨基酰化 O-tRNA 的例如小于约 10%、小于约 5%或小于约 1%的效率。

具体实施方式

适于制造包括一种或一种以上例如非天然氨基酸的所选氨基酸的蛋白质的转译系统，描述于标题为“METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL tRNA SYNTHETASE PAIRS”的美国专利申请案 10/126,931 和标题为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS”的美国专利申请案 10/126,927 中。另外，参见标题为“EXPANDING THE EUKARYOTIC GENETIC CODE” US 2002/01825,867。每一所述申请案都以引用的方式全部并入本文中。所述转译系统通常包含，包括正交 tRNA (O-tRNA)、正交氨基酰基 tRNA 合成酶 (O-RS) 和例如非天然氨基酸的所选氨基酸的细胞，其中 O-RS 用所选氨基酸氨基酰化 O-tRNA。本发明的正交对包含 O-tRNA (例如抑制基因 tRNA、移码 tRNA 等等) 和 O-RS。O-tRNA 识别第一选择密码子并且在同源合成酶的存在下响应选择密码子而具有抑制活性。细胞使用所述组分将所选氨基酸并入生长多肽链中。举例而言，包含编码所关注的多肽的多核苷酸的核酸也可存在，其中多核苷酸包含通过 O-tRNA 识别出的选择密码子。转译系统也可作为活体外系统。本发明的 RS 分子适用于包括在转译中利用核糖体的系统中的任何转译系统。

转译系统也可以是无细胞（活体外）转译系统。在可包括 mRNA 作为模板（活体外转译）或者 DNA 作为模板（经组合的活体外转录和转译）的所述系统中，活体外合成通过核糖体引导。已经对开发无细胞蛋白质表达系统进行了相当大的努力。参见（例如），Kim, D.M.和 J.R. Swartz, *Biotechnology and Bioengineering*, 74 :309-316 (2001); Kim, D.M.和 J.R. Swartz, *Biotechnology Letters*, 22, 1537-1542, (2000); Kim, D.M.和 J.R. Swartz, *Biotechnology Progress*, 16, 385-390, (2000); Kim, D.M.和 J.R. Swartz, *Biotechnology and Bioengineering*, 66, 180-188, (1999); 和 Patnaik, R.和 J.R. Swartz, *Biotechniques* 24, 862-868, (1998); 美国专利第 6,337,191 号; 美国专利公开案第 2002/0081660 号; WO

00/55353; WO 90/05785, 所述文献是以引用的方式并入本文中。可以应用的另一种方法包括 mRNA-肽融合技术。参见(例如), R. Roberts 和 J. Szostak, *Proc. Natl Acad. Sci. (USA)* 94:12297-12302 (1997); A. Frankel 等人, *Chemistry & Biology* 10:1043-1050 (2003)。在所述方法中, 与嘌呤霉素连接的 mRNA 模板在核糖体上被转译成肽。如果已经修饰一个或一个以上 tRNA 分子, 那么也可将非天然氨基酸并入肽中。在已经阅读末尾 mRNA 密码子后, 嘌呤霉素俘获肽的 C-末端。如果在活体外检定中发现所得 mRNA-肽结合物具有所关注的性质, 那么其同一性可容易地从 mRNA 序列展现。这样, 所属领域的技术人员可以筛选包含一种或一种以上非天然编码的氨基酸的多肽的库, 以鉴别具有所要性质的多肽。近期, 已经报道用经纯化的组分的活体外核糖体转译允许合成经非天然编码的氨基酸取代的肽。参见(例如) A. Forster 等人, *Proc. Natl Acad. Sci. (USA)* 100:6353 (2003)。

在某些实施例中, 包含本发明的 RS 的大肠杆菌细胞包括所述转译系统。举例而言, 本发明的大肠杆菌细胞包括正交 tRNA (O-tRNA), 其中 O-tRNA 包含在同源合成酶存在下响应选择密码子的抑制活性; 正交氨酰基-tRNA 合成酶 (O-RS); 所选氨基酸; 和包含编码所关注的多肽的多核苷酸的核酸, 其中多核苷酸包含由 O-tRNA 识别的选择密码子。

本发明也提供在容许并入一种以上所选氨基酸的细胞中的多个 O-tRNA/O-RS 对。在某些实施例中, 细胞可另外包括其他不同的 O-tRNA/O-RS 对和第二所选氨基酸, 其中 O-tRNA 识别第二选择密码子并且 O-RS 用第二所选氨基酸优先氨酰化 O-tRNA。举例而言, 细胞可另外包含(例如)由詹氏甲烷球菌的酪氨酰基 tRNA 合成酶得到的琥珀抑制基因 tRNA-氨酰基 tRNA 合成酶对。

O-tRNA 和/或 O-RS 可为天然存在的或可通过来自各种生物体的天然存在的 tRNA 和/或 RS 的突变而得到, 例如所述突变产生 tRNA 的库和/或 RS 的库。举例而言, 一种生产正交 tRNA/氨酰基-tRNA 合成酶对的策略涉及将来自(例如)不同于宿主细胞的来源或多个来源的异源性 tRNA/合成酶对引进宿主细胞中。异源性合成酶候选者的性质包括,(例如)其不装填任何宿主细胞 tRNA, 并且异源性 tRNA 候选者的性质包括,(例如)其不通过任何宿主细胞合成酶而氨酰化。另外, 异源性 tRNA 与所有宿主细胞合成酶正交。

用于产生正交对的第二种策略涉及产生从其筛选和/或选择 O-tRNA 或 O-RS 的突变体库。所述策略也可组合。

在各种实施例中, O-tRNA 和 O-RS 是由至少一种生物体得到。在另一个实施例中,

O-tRNA 是由来自第一生物体的天然存在的或经突变的天然存在的 tRNA 得到，并且 O-RS 是由来自第二生物体的天然存在的或经突变的天然存在的 RS 得到。在一个实施例中，第一生物体和第二生物体是不同的。举例而言，正交对可以包括由横川病毒得到的 tRNA 合成酶和由古细菌 tRNA（例如来自盐杆菌种 *NRC-1*）得到的 tRNA。或者，第一生物体和第二生物体是相同的。为补充信息，参见本文中标题为“来源和宿主生物体”的部分。

在本发明的某些实施例中，本发明的 O-RS 包含如 SEQ ID NO: 4 中所述的多核苷酸序列或其互补多核苷酸序列或其保守变化，或通过如 SEQ ID NO: 4 中所述的多核苷酸序列或其互补多核苷酸序列或其保守变化来编码。在某些实施例中，O-RS 包含如 SEQ ID NO: 5 中所述的氨基酸序列。又参见本文中标题为“核酸和多肽序列和变体”的部分。

正交 tRNA (O-tRNA)

正交 tRNA (O-tRNA) 介导将所选氨基酸（例如）活体内或活体外并入蛋白质中，所述蛋白质通过包含由 O-tRNA 识别的选择密码子的多核苷酸来编码。O-tRNA 可以通过包括（但不限于）化学氨基酰化或酶促氨基酰化的任何方法或技术用所要氨基酸来氨基酰化。经氨基酰化的 O-tRNA 可以被直接地添加到转译系统中。O-tRNA 可以通过本发明的 RS 用所选氨基酸活体外或活体内氨基酰化。另外，RS 可为 O-RS。O-tRNA 可直接地或通过提供编码 O-tRNA 或其部分的多核苷酸而提供到转译系统（例如，活体外转译组分或细胞）中。举例而言，O-tRNA 或其部分通过如 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3 中所述的多核苷酸序列或其互补多核苷酸序列或其保守变化来编码。O-RS 可直接地（例如 SEQ ID NO: 5 或 SEQ ID NO: 17 或其保守变化）或通过提供编码 O-RS 或其部分的多核苷酸而提供到转译系统（例如，活体外转译组分或细胞）中。举例而言，O-RS 或其部分通过如 SEQ ID NO: 4 中所述的多核苷酸序列、编码氨基酸序列 SEQ ID NO: 17 的多核苷酸序列或其互补多核苷酸序列或其保守变化来编码。

本发明的 O-tRNA 包含在同源合成酶存在下响应选择密码子的抑制活性。抑制活性可通过所属领域中已知的多种检定中的任何一种来测定。举例而言，可使用 β -半乳糖苷酶报告基因检定。（例如）在 *lacZ* 的肽 VVLQRRDWEN 的 Leu-25 经例如 TAG、TGA、AGGA 等密码子的选择密码子或酪氨酸、丝氨酸、白氨酸等的有义密码子（作为对照）置换时，使用在启动子控制下表达 *lacZ* 基因的质粒的衍生物。将经衍生的 *lacZ* 质粒连同包含本发明的 O-tRNA 的质粒一起引入来自适当生物体（例如，其中可使用正交组分的生物体）的细胞中。也可引入同源合成酶（作为多肽或者当被表达时编码同源合成酶的多核苷酸）。使细胞在培养基中生长到所要密度，例如生长到 OD₆₀₀ 为约 0.5，并且（例

如)使用 BetaFluor™ β -半乳糖苷酶检定试剂盒 (Novagen) 执行 β -半乳糖苷酶检定。抑制百分比是按样本相对于例如从经衍生的 lacZ 构筑体观察到的值的可比较对照物的活性的百分比来计算, 其中构筑体在所要位置具有对应的有义密码子而不是选择密码子。

适用于本发明的 O-tRNA 的实例是揭示于美国专利申请案 10/126,931、10/126,927 和 10/825,867 中的 O-tRNA 分子的任何一种分子。在 tRNA 分子中, 胸苷嘧啶 (T) 经尿嘧啶 (U) 置换。另外, 可存在对碱基的其他修饰。本发明也包括 O-tRNA 的保守变化。举例而言, O-tRNA 的保守变化包括如 O-tRNA 一样起作用并且维持 tRNA L-形结构, 但不具有相同序列 (并且不同于野生型 tRNA 分子) 的分子。又参见本文中标题为“核酸和多肽序列和变体”的部分。

包含 O-tRNA 的组合物可另外包括正交氨酰基-tRNA 合成酶 (O-RS), 其中 O-RS 用所选氨基酸 (例如非天然氨基酸) 优先氨酰化 O-tRNA。在某些实施例中, 包括 O-tRNA 的组合物可另外包括转译系统 (例如, 活体外或活体内转译系统)。包含编码所关注的多肽的多核苷酸的核酸也可存在于细胞或其他转译系统中, 其中所述多核苷酸包含一个或一个以上由 O-tRNA 识别的选择密码子, 或所述选择密码子的一个或一个以上的组合。又参见, 本文中标题为“正交氨酰基-tRNA 合成酶 (O-RS)”的部分。

生产例如 O-tRNA 的正交 tRNA (O-tRNA) 的方法也是本发明的特征。通过所述方法生产的例如 O-tRNA 的 tRNA 也是本发明的特征。

生产正交 tRNA 的方法包括使 tRNA 的集合中的每一 tRNA 的反密码子环突变以容许识别选择密码子 (例如, 琥珀密码子、蛋白石密码子、4 碱基密码子等), 进而提供多个可能的 O-tRNA; 且分析多个可能的 O-tRNA 的成员的二级结构以鉴别二级结构中的非规范碱基对; 且视需要使非规范碱基对突变 (例如, 使非规范碱基对突变成规范碱基对)。非规范碱基对可位于二级结构的茎区。O-tRNA 可以具有一种或一种以上特征或活性的改善, 诸如与例如多个 tRNA 序列的起始物质相比, 所要生物体的正交性的改善, 同时保持其对所要 RS 的亲合力。

或者, O-tRNA 可以通过使已知 tRNA 突变来发展以调节其与一种或一种以上影响转译或为转译机器的组分的分子的相互作用或对所述分子的结合亲和力。所述组分包括 (但不限于) 延伸因子。细菌延伸因子 EF-Tu 在蛋白质合成的延伸步骤中起关键作用。在 tRNA 通过 tRNA 合成酶氨酰化后, EF-Tu 结合经氨酰化的 tRNA 并且将其带到核糖体的 A 部位。由于 EF-Tu 与经氨酰化的 tRNA 之间的结合, 经装填的氨基酸与 tRNA 之间的酯键受自发水解作用的保护。Stortchevoi 等人研究了在 T^{YC} 茎中的大肠杆菌起始 tRNA^{fMet} U50:G64 摆动碱基对的突变体, 因为发现所述碱基对为阻断延伸中的

tRNA 活性的第二负性决定子，估计可能由于 EF-Tu.GTP 与经氨基酰化的 tRNA 之间减弱的相互作用 (JBC 2003 278 (20) :17672-17679)。同样，LaRiviere 等人在 Science 2001 年 10 月 5 日;294 (5540) : 165-8 中描述了氨基酸和 tRNA 体对与 EF-Tu 的总结合亲和力的热力学贡献。他们指出，tRNA 体和氨基酸的贡献彼此独立并且当 tRNA 经正确酰化时，其相互补充。对 EF-Tu.GTP 与经非天然氨基酸氨基酰化的 tRNA 之间相互作用的改变可以影响将 tRNA 加载到核糖体的 A 部位的效率。可能的突变部位也可以通过分析 tRNA 与诸如 EF-Tu 的转译机器的其他组分之间的复合体的晶体结构而发现。举例而言，Nissen 等人已指出，EF-Tu.GTP 与酵母苯丙氨酰基-转移 RNA (Phe-tRNA) 的 T Ψ C 茎的磷酸酯骨架直接结合 (Science 1995 270 (5241) :1464-1472)。

所述方法视需要包括分析 tRNA 和/或氨酰基-tRNA 合成酶的序列同源性，以确定似乎对特定生物体来说为正交的 O-tRNA、O-RS 和/或其对的可能候选者。可使用所属领域中已知的和本文中描述的计算机程序进行分析。在一个实例中，为选择适用于原核生物体的可能的正交转译组分，选择对原核生物体不显示异常同源性的合成酶和/或 tRNA。

tRNA 的集合也可通过一致的策略生产。举例而言，tRNA 的集合是通过比对多个 tRNA 序列；确定一致序列；和使用至少一部分、大部分或全部一致序列产生 tRNA 的库而生产。举例而言，一致序列可用例如 GCG 程序堆积 (GCG program *pileup*) 的计算机程序来编辑。视需要，将通过所述程序确定的简并位置 (degenerate position) 变成在所述位置处最常见的碱基。库通过所属领域中已知的技术使用一致序列合成。举例而言，寡核苷酸的重叠延伸作用 (其中 tRNA 基因的各部位可被合成为 90%一致序列和 10%其他 3 个碱基的混合物的掺杂混合物) 可用以提供基于一致序列的库。也可使用其他混合物，例如 75%一致序列和 25%其他 3 个碱基的混合物、80%一致序列和 20%其他 3 个碱基的混合物、95%一致序列和 5%其他 3 个碱基的混合物等。

突变 tRNA 的库可使用所属领域中已知的各种突变技术产生。举例而言，突变 tRNA 可通过部位特异性突变、随机点突变、同源重组、DNA 滑移或其他递归突变方法、嵌合建构或其任何组合产生。

其他突变可在例如反密码子环、接纳茎 (acceptor stem)、D 臂或环、可变环、T Ψ C 臂或环的 tRNA 的所要环或区域、tRNA 分子的其他区域或其组合中的例如非保守位置或保守位置、随机化位置或其组合的特定位置上引入。突变可以包括在茎区中的匹配碱基对。

通常，O-tRNA 通过使第一物种的细胞群体经受负性选择而获得，其中细胞包含多个可能的 O-tRNA 的成员。负性选择消除包含多个可能的 O-tRNA 的成员的细胞，所述

O-tRNA 通过对细胞来说为内源性的氨酰基-tRNA 合成酶 (RS) 氨基酰化。所述负性选择提供与第一物种的细胞正交的 tRNA 的集合。

在负性选择的某些实施例中, 将选择密码子引入多核苷酸中, 所述多核苷酸编码负性选择标记物, 例如给与抗生素抗性的酶, 例如 β -内酰胺酶; 给与可检测产物的酶, 例如 β 半乳糖苷酶; 氯霉素乙酰转移酶 (CAT), 例如在非基本位置处的毒性产物, 诸如 barnase; 等。筛选/选择可通过在选择剂 (例如, 抗生素, 诸如氨苄青霉素 (ampicillin)) 存在下, 使细胞群体生长来进行。在一个实施例中, 选择剂的浓度是变化的。

举例而言, 为测量抑制基因 tRNA 的活性, 使用基于例如无义密码子的选择密码子的活体内抑制, 或引入编码例如 β -内酰胺酶 (*bla*) 的基因的负性选择标记物的多核苷酸中的移码突变的选择系统。举例而言, 建构在某个位置上具有 (例如) TAG、AGGA 和 TGA 的多核苷酸变体, 例如 *bla* 变体。例如细菌的细胞经所述多核苷酸转形。在不能通过内源性大肠杆菌合成酶有效地装填的正交 tRNA 的情况下, 例如氨苄青霉素抗性的抗生素抗性应约为或小于不经质粒转形的细菌的抗生素抗性。如果 tRNA 是非正交的, 或如果能够装填 tRNA 的异源性合成酶在系统中被共表达, 那么观察到高水平的例如氨苄青霉素的抗生素的抗性。选择在抗生素浓度约等于不经质粒转形的细胞的 LB 琼脂培养盘上不能生长的细胞, 例如细菌。

在毒性产物 (例如核糖核酸酶 barnase) 的情况下, 当多个可能的 tRNA 的成员通过例如大肠杆菌合成酶的内源性宿主氨基酰化 (也就是说, 其不与例如大肠杆菌合成酶的宿主正交) 时, 选择密码子被抑制并且所产生的毒性多核苷酸产物导致细胞死亡。怀有正交 tRNA 或无功能的 tRNA 的细胞存活。

在一个实施例中, 随后使与所要生物体正交的 tRNA 的集合经受正性选择, 其中将选择密码子放置在 (例如) 通过诸如 β -内酰胺酶基因的耐药性基因编码的正性选择标记物中。正性选择是在包含编码或包含 tRNA 的集合的成員的多核苷酸、编码正性选择标记物的多核苷酸和编码同源 RS 的多核苷酸的细胞上执行。所述多核苷酸在细胞中表达并且细胞在例如氨苄青霉素的选择剂存在下生长。然后, 就响应所述选择密码子 tRNA 通过共表达的同源合成酶氨基酰化的能力和插入氨基酸的能力来选择 tRNA。通常, 与怀有无功能的 tRNA 或不能由所关注的合成酶有效识别出的 tRNA 的细胞相比, 所述细胞展示抑制效率的增强。怀有无功能 tRNA 或通过所关注的合成酶不有效识别出的 tRNA 的细胞对抗生素敏感。因此, (i) 不为内源性宿主 (例如大肠杆菌) 合成酶的底物; (ii) 可通过所关注的合成酶氨基酰化; 和 (iii) 在转译中起作用的 tRNA 在两种选择中存活。

上述方法中选择 (例如正性选择、负性选择或正性和负性选择) 的严格性视需要可

以变化。举例而言，因为 barnase 是极具毒性的蛋白质，所以负性选择的严格性可通过将不同数量的选择密码子引入 barnase 基因中和/或通过使用诱导型启动子而控制。在另一个实例中，选择或筛选剂（例如氨苄青霉素）的浓度是变化的。在一个方面中，因为在早期循环期间，所要活性可为低的，所以严格性是变化的。因此，在早期循环中应用较低严格性的选择准则并且在选择的后期循环中应用更严格的准则。在某些实施例中，负性选择、正性选择或负性和正性选择可重复多次。可使用多种不同的负性选择标记物、正性选择标记物或负性和正性选择标记物。在某些实施例中，正性和负性选择标记物可为相同的。

其他类型的选择/筛选可用于本发明以生产例如 O-tRNA、O-RS 和 O-tRNA/O-RS 对的正交转译组分。举例而言，负性选择标记物、正性选择标记物或正性和负性选择标记物可包括发荧光或在适合的反应物存在下催化发光反应的标记物。在另一个实施例中，标记物的产物通过荧光活化细胞拣选（FACS）或通过发光来检测。视需要，标记物包括基于亲和力的筛选标记物。参见，Francisco, J. A. 等人，(1993) *Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface*. Proc Natl Acad Sci U S A. 90:10444-8。

生产重组正交 tRNA 的其他方法可见于（例如）标题为“Methods and Compositions for the Production of Orthogonal tRNA-Aminoacyl tRNA Synthetase Pairs”的美国专利申请案 10/126,931 和标题为“In vivo Incorporation of Unnatural Amino Acids”的美国专利申请案 10/126,127, 和标题为“EXPANDING THE EUKARYOTIC GENETIC CODE”的 USSN 10/825,867 中。又参见 Forster 等人，(2003) *Programming peptidomimetic synthetases by translating genetic codes designed de novo* PNAS 100(11):6353-6357; 和 Feng 等人，(2003), *Expanding tRNA recognition of a tRNA synthetase by a single amino acid change*, PNAS 100(10): 5676-5681。

tRNA 可以通过包括（但不限于）化学氨基酰化或酶促氨基酰化的任何方法或技术用所要氨基酸氨基酰化。

氨基酰化可以通过氨酰基 tRNA 合成酶或通过包括（但不限于）核糖酶的其他酶促分子完成。术语“核糖酶”可与“催化性 RNA”互换。Cech 和同事（Cech, 1987, *Science*, 236:1532-1539; McCorkle 等人，1987, *Concepts Biochem.* 64:221-226）证明可担当催化剂（核糖酶）的天然存在的 RNA 的存在。然而，尽管已经展示所述天然 RNA 触媒仅对用于裂解和剪接的核糖核酸底物起作用，但是核糖酶的人工进化的近期发展已经将催化作用的清单扩展到各种化学反应。研究已经鉴别可在其自身(2') 3'-末端上催化氨酰基

-RNA 键的 RNA 分子 (Illangakekare 等人, 1995 Science 267:643-647) 和可将氨基酸从一个 RNA 分子转移到另一个 RNA 分子的 RNA 分子 (Lohse 等人, 1996, Nature 381:442-444)。

以引用的方式并入本文中的美国专利申请公开案 2003/0228593, 描述建构核糖酶的方法和其在用天然编码的和非天然编码的氨基酸氨基酰化 tRNA 中的用途。包括 (但不限于) 核糖酶的可氨基酰化 tRNA 的酶促分子的底物固定形式, 可以使经氨基酰化的产物能够有效亲和力纯化。适合底物的实例包括琼脂糖、琼脂糖凝胶和磁性珠粒。用于氨基酰化的核糖酶的底物固定形式的生产和用途描述于 Chemistry and Biology 2003, 10:1077-1084 和美国专利申请公开案 2003/0228593 中, 其都以引用的方式并入本文中。

化学氨基酰化方法包括 (但不限于) 由 Hecht 和同事 (Hecht, S. M. Acc. Chem. Res. 1992, 25, 545; Heckler, T. G.; Roesser, J. R.; Xu, C.; Chang, P.; Hecht, S. M. Biochemistry 1988, 27, 7254; Hecht, S. M.; Alford, B. L.; Kuroda, Y.; Kitano, S. J. Biol. Chem. 1978, 253, 4517) 和由 Schultz、Chamberlin、Dougherty 和其他人 (Cornish, V. W.; Mendel, D.; Schultz, P. G. Angew. Chem. Int. Ed. Engl. 1995, 34, 621; Robertson, S. A.; Ellman, J. A.; Schultz, P. G. J. Am. Chem. Soc. 1991, 113, 2722; Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. Science 1989, 244, 182; Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R. J. Am. Chem. Soc. 1989, 111, 8013; Bain, J. D.等人 Nature 1992, 356, 537; Gallivan, J. P.; Lester, H. A.; Dougherty, D. A. Chem. Biol. 1997, 4, 740; Turcatti 等人, J. Biol. Chem. 1996, 271, 19991; Nowak, M. W.等人, Science, 1995, 268, 439; Saks, M. E.等人 J. Biol. Chem. 1996, 271, 23169; Hohsaka, T.等人 J. Am. Chem. Soc. 1999, 121, 34) 所介绍的避免在氨基酰化中使用合成酶的方法。所述方法或其他化学氨基酰化方法可用以氨基酰化 tRNA 分子。

使用经化学修饰的氨酰基-tRNA 的生物合成方法已经用以将若干生物物理学探针活体外并入所合成的蛋白质中。参见以下公开案和其中引用的参考文献: Brunner, J. *New Photolabeling and crosslinking methods*, Annu. Rev Biochem, 62:483-514 (1993); 和 Krieg, U.C., Walter, P., Hohnson, A.E. *Photocrosslinking of the signal sequence of nascent preprolactin of the 54-kilodalton polypeptide of the signal recognition particle*, Proc. Natl. Acad. Sci. 83 (22): 8604-8608 (1986)。

以前, 已经展示, 非天然氨基酸可通过将经化学氨基酰化的抑制基因 tRNA 添加到用含有所要琥珀无义突变的基因程式化的蛋白质合成反应中而在部位特异地活体外并入蛋白质中。使用所述方法, 所属领域的技术人员可用封闭结构的同源物取代大量常见

的 20 种氨基酸，例如，氟苯丙氨酸取代苯丙氨酸，使用营养缺陷菌株取代具体的氨基酸。参见（例如），Noren, C.J., Anthony-Cahill, Griffith, M.C., Schultz, P.G. *A general method for site-specific incorporation of unnatural amino acids into proteins*, Science, 244: 182-188 (1989); M.W. Nowak 等人, Science 268:439-42 (1995); Bain, J.D., Glabe, C.G., Dix, T.A., Chamberlin, A.R., Diala, E.S. *Biosynthetic site-specific Incorporation of a non-natural amino acid into a polypeptide*, J. Am Chem Soc, 111:8013-8014 (1989); N. Budisa 等人, FASEB J. 13:41-51 (1999); Ellman, J.A., Mendel, D., Anthony-Cahill, S., Noren, C.J., Schultz, P.G. *Biosynthetic method for introducing unnatural amino acids site-specifically into proteins*. Methods in Enz., 第 202 卷, 301-336 (1992); 和 Mendel, D., Cornish, V.W.和 Schultz, P.G. *Site-Directed Mutagenesis with an Expanded Genetic Code*, Annu Rev Biophys. Biomol Struct. 24,435-62(1995)。

举例而言，制备识别终止密码子 UAG 的抑制基因 tRNA 并且用非天然氨基酸将其化学氨基酰化。惯用的定位突变用以在蛋白质基因中的所关注的部位引入终止密码子 TAG。参见（例如），Sayers, J.R., Schmidt, W. Eckstein, F. *5'-3' Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis*, Nucleic Acids Res. 16(3):791-802 (1988)。当经酰化的抑制基因 tRNA 和突变基因在活体外转录/转译系统中组合时，响应 UAG 密码子将非天然氨基酸并入，得到在规定位置上含有那个氨基酸的蛋白质。使用 [³H]-Phe 的实验和用 α -羧酸的实验证明，仅仅所要氨基酸在规定的位点上通过 UAG 密码子并入，并且所述氨基酸未在蛋白质中的任何其他部位并入。参见（例如），Noren 等人，同上；Kobayashi 等人，(2003) *Nature Structural Biology* 10(6):425-432；和 Ellman, J.A., Mendel, D., Schultz, P.G. *Site-specific incorporation of novel backbone structures into proteins*, Science, 255(5041): 197-200 (1992)。

用于产生催化性 RNA 的方法可以涉及产生随机化的核糖酶序列的分离池、对池执行定位进化、为所需氨基酰化活性而筛选池和选择显示所要氨基酰化活性的所述核糖酶的序列。

核糖酶可包含有利于酰化活性的基元和/或区域，诸如 GGU 基元和 U 富集区域。举例而言，已经报道 U 富集区域可促进氨基酸底物的识别，并且 GGU-基元可与 tRNA 的 3'末端形成碱基对。在组合时，GGU 和基元和 U 富集区域同时促进氨基酸和 tRNA 的同时识别，并且进而促进 tRNA 的 3'末端的氨基酰化。

核糖酶可通过使用与 tRNA^{Asn}_{CCCG} 结合的部分随机化的 r24mini 进行活体外选择，接着通过见于活性无性系中的一致序列的系统工程化而产生。通过所述方法获得的示范

性核糖酶被称为“Fx3 核糖酶”并且描述于美国公开申请案第 2003/0228593 号中，所述专利的内容是以引用的方式并入本文中，所述核糖酶担当多用途催化剂用于合成经同源非天然氨基酸装填的各种氨酰基-tRNA。

底物上的固定可用以促成经氨酰化的 tRNA 的有效亲和力纯化。适合的底物的实例包括（但不限于）琼脂糖、琼脂糖凝胶和磁性珠粒。核糖酶可通过利用 RNA 的化学结构固定于树脂上，诸如 RNA 的核糖上的 3'-顺式二醇可经高碘酸盐氧化以产生对应的二醛以促进 RNA 在树脂上的固定。可使用包括廉价酰肼树脂的各种类型的树脂，其中还原性胺化作用使树脂与核糖酶之间的相互作用成为不可逆的键。氨酰基-tRNA 的合成可通过所述管柱上氨酰化技术而显著地促进。Kourouklis 等人 *Methods* 2005; 36:239-4 描述基于管柱的氨酰化系统。

经氨酰化的 tRNA 的分离可以各种方式完成。一种适合的方法是，从具有诸如含有 10 mM EDTA 的乙酸钠溶液的缓冲液、含有 50 mM N-(2-羟乙基)哌嗪-N'-(3-丙烷磺酸)、pH 7.0 的 12.5 mM KCl、10 mM EDTA 的缓冲液或仅为经 EDTA 缓冲的水（pH 7.0）的管柱洗提经氨酰化的 tRNA。

可将经氨酰化的 tRNA 添加到转译反应中，以便并入氨基酸，在通过转译反应制得的多肽中的特别位置上，tRNA 被所述氨基酸氨酰化。可以使用经氨酰化的 tRNA 的转译系统的实例，包括（但不限于）细胞溶胞产物。细胞溶胞产物提供为从输入 mRNA 活体外转译多肽所必需的反应组分。所述反应组分的实例包括（但不限于）核糖体蛋白、rRNA、氨基酸、tRNA、GTP、ATP、转译起始因子和延伸因子和与转译相关的其他因子。另外，转译系统可以是分批转译或隔开转译。分批转译系统在单一隔室中组合反应组分，而隔开转译系统使转译反应组分与可抑制转译效率的反应产物分离。所述转译系统是市售的。

另外，可以使用偶合转录/转译系统。偶合转录/转译系统允许将输入的 DNA 转录成对应的 mRNA，而 mRNA 又被反应组分转译。市售偶合转录/转译的实例是 Rapid Translation System (RTS, Roche Inc.)。系统包括含有用于提供诸如核糖体的转译组分的大肠杆菌溶胞产物和转译因子的混合物。另外，包括 RNA 聚合酶用于将输入 DNA 转录成适用于转译的 mRNA 模板。RTS 可通过插入反应隔室（包括供应/消耗隔室和转录/转译隔室）之间的膜使反应组分隔开。

tRNA 的氨酰化可以通过包括（但不限于）移转酶、聚合酶、催化性抗体、多官能蛋白质，诸如此类的其他试剂执行。

正交氨酰基-tRNA 合成酶 (O-RS)

本发明的 O-RS 优先地用所选氨基酸在活体外或活体内氨基酰化 O-tRNA。本发明的 O-RS 可通过包括 O-RS 的多肽和/或通过编码 O-RS 或其部分的多核苷酸而提供到转译系统（例如，活体外转译组分或细胞）中。举例而言，O-RS 或其部分通过如 SEQ ID NO. : 4 中所述的多核苷酸序列或其互补多核苷酸序列或其保守变化编码。本发明的 O-RS 可以氨基酰化大量不同的 O-tRNA 分子，包括（但不限于）本文中揭示的所述 O-tRNA 分子。

用于鉴别供 O-tRNA（例如 O-tRNA）使用的正交氨酰基-tRNA 合成酶（O-RS）（例如 O-RS）的方法也是本发明的特征。举例而言，方法包括使第一物种的细胞群体经受正性选择，其中所述细胞各自包含：1）多个氨酰基-tRNA 合成酶（RS）的成员，其中多个 RS 包含突变体 RS、由不同于第一物种的物种得到的 RS 或突变体 RS 和由不同于第一物种的物种得到的 RS 两者；2）来自第二物种的正交 tRNA（O-tRNA）；和 3）编码正性选择标记物并且包含至少一个选择密码子的多核苷酸。对于与缺乏多个 RS 的成员或具有降低量的多个 RS 的成员的细胞相比，展示抑制效率增强的所述细胞来选择或筛选细胞。具有抑制效率增强的细胞包含氨基酰化 O-tRNA 的活性 RS。将来自第一物种的第一组 tRNA 的通过活性 RS 氨基酰化（活体外或活体内）的水平，与来自第二物种的第二组 tRNA 的通过活性 RS 氨基酰化（活体外或活体内）的水平相比较。氨基酰化的水平可通过可检测物质（例如，经标记的氨基酸或非天然氨基酸）测定。选择与第一组 tRNA 相比，更有效地氨基酰化第二组 tRNA 的活性 RS，进而提供供 O-tRNA 使用的正交氨酰基-tRNA 合成酶。通过所述方法鉴别的例如 O-RS 的 O-RS 也是本发明的特征。

可使用大量检定中的任何检定来测定氨基酰化。所述检定可在活体外或活体内执行。举例而言，活体外氨基酰化检定描述于（例如）Hoben, P.和 Soil, D. (1985) Methods Enzymol. 113:55-59 和美国专利申请公开案第 2003/0228593 号中。氨基酰化也可通过使用报告基因以及正交转译组分，并且检测在细胞中表达包含至少一个编码蛋白质的选择密码子的多核苷酸的报告基因而测定。又参见，标题为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS”的美国专利申请案 10/126,927 和标题为“EXPANDING THE EUKARYOTIC GENETIC CODE”的 USSN 10/825,867。

经鉴别的 O-RS 可另外经操纵以改变合成酶的底物特异性以便仅所要非天然氨基酸，而不是常见 20 种氨基酸的任何氨基酸被装填到 O-tRNA 中。产生对非天然氨基酸具有底物特异性的正交氨酰基 tRNA 合成酶的方法包括，使合成酶（例如）在合成酶中的活性部位、在合成酶中的编辑机制部位、在通过组合合成酶的不同域的不同部位等处突变，和应用选择处理。使用基于正性选择接着负性选择的组合的策略。在正性选择中，

在正性标记物的非基本位置引入的选择密码子的抑制作用使细胞在正性选择压力下存活。在天然氨基酸和非天然氨基酸存在下，存活性因此编码用天然氨基酸或非天然氨基酸装填正交抑制基因 tRNA 的活性合成酶。在负性选择中，在负性标记物的非基本位置引入的选择密码子的抑制作用除去具有天然氨基酸特异性的合成酶。负性和正性选择的存活性编码仅用非天然氨基酸氨基酰化（装填）正交抑制基因 tRNA 的合成酶。所述合成酶可随后经受例如 DNA 滑移或其他递归突变方法的其他突变。

突变体 O-RS 的库可使用所属领域中已知的各种突变技术产生。举例而言，突变体 RS 可通过部位特异性突变、随机点突变、同源重组、DNA 滑移或其他递归突变方法、嵌合建构或其任何组合产生。举例而言，突变体 RS 的库可由两个或两个以上的其他（例如更小、较少变化）的“子库”产生。RS 的嵌合库也包括在本发明中。应注意，来自各种生物体（例如，诸如真细菌或太古细菌的微生物）的 tRNA 合成酶的库，诸如包含天然多样性的库（参见（例如），Short 等人的美国专利第 6,238,884 号；Schallenberger 等人的美国专利第 5,756,316 号；Petersen 等人的美国专利第 5,783,431 号；Thompson 等人的美国专利第 5,824,485 号；Short 等人的美国专利第 5,958,672 号）视需要经建构且对于正交对筛选。

一旦合成酶经受正性和负性选择/筛选策略，所述合成酶就可随后经受进一步突变。举例而言，可分离编码 O-RS 的核酸；可由所述核酸产生一组编码经突变的 O-RS（例如，通过随机突变、部位特异性突变、重组或其任何组合）的多核苷酸；并且，可重复所述单独步骤或所述步骤的组合直到获得用非天然氨基酸优先氨基酰化 O-tRNA 的经突变的 O-RS。在本发明的一个方面中，将所述步骤执行多次，例如至少 2 次。

其他水平的选择/筛选严格性也可用于本发明的方法中，以生产 O-tRNA、O-RS 或其对。选择或筛选严格性可在用以生产 O-RS 的方法的一个或两个步骤中变化。其可包括（例如）改变所使用的选择/筛选剂的量等。也可执行正性和/或负性选择的其他循环。选择或筛选也可包含一次或多次包括（例如）氨基酸通透性的改变、转译效率的改变、转译保真度的改变等的正性或负性选择或筛选。通常，所述一种或一种以上改变是基于其中使用正交 tRNA-tRNA 合成酶对来生产蛋白质的生物体中的一个或一个以上基因的突变。

其他类型的选择可用于本发明中，以用于（例如）O-RS、O-tRNA 和 O-tRNA/O-RS 对。正性选择标记物可为包括（但不限于）为生长提供营养补充的产物的各种分子中的任何分子，并且选择是在缺乏营养补充的培养基上执行。编码正性选择标记物的多核苷酸的实例包括（但不限于）（例如）基于补充细胞的营养缺陷氨基酸的报告基因、his3

基因（例如，其中通过提供 3-氨基三唑（3-AT）所检测，his3 基因编码咪唑磷酸甘油脱水酶）、ura3 基因、leu2 基因、lys2 基因、lacZ 基因、adh 基因等。参见（例如），G. M. Kishore 和 D. M. Shah, (1988), Amino acid biosynthesis inhibitors as herbicides, *Annual Review of Biochemistry* 57:627-663。在一个实施例中，lacZ 生产是通过邻硝基苯基- β -D-吡喃半乳糖苷（ONPG）水解作用检测。参见（例如），I. G. Serebriiskii 和 E. A. Golemis, (2000), Uses of lacZ to study gene function: evaluation of beta-galactosidase assays employed in the yeast two-hybrid system, *Analytical Biochemistry* 285:1-15。其他正性选择标记物包括（例如）荧光素酶、绿色荧光蛋白质（GFP）、YFP、EGFP、RFP、抗生素抗性基因的产物（例如，氯霉素乙酰转移酶（CAT））、转录调节剂蛋白质（例如 GAL4）等。视需要，编码正性选择标记物的多核苷酸包含选择密码子。

编码正性选择标记物的多核苷酸可操作性地与反应元素连接。也可存在编码调节从反应元素的转录并且包含至少一个选择密码子的转录调节剂蛋白质的其他多核苷酸。通过经非天然氨基酸氨基酰化的 O-tRNA 将非天然氨基酸并入转录调节剂蛋白质中，产生编码正性选择标记物的多核苷酸（例如报告基因）的转录。视需要，选择密码子位于或大体上靠近编码转录调节剂蛋白质的 DNA 结合域的多核苷酸的一部分。

编码负性选择标记物的多核苷酸也可操作性地与反应元素连接，从所述反应元素的转录是通过转录调节剂蛋白质介导。参见（例如），A. J. DeMaggio 等人, (2000), The yeast split-hybrid system, *Method Enzymol.* 328:128-137; H. M. Shih 等人, (1996), A positive genetic selection for disrupting protein-protein interactions: identification of CREB mutations that prevent association with the coactivator CBP, *Proc. Natl. Acad. Sci. U. S. A.* 93:13896-13901; M. Vidal 等人, (1996), Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system, *Proc. Natl. Acad. Sci. U. S. A.* 93:10321-10326; 和 M. Vidal 等人, (1996), Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions (*Proc. Natl. Acad. Sci. U. S. A.* 93:10315-10320)。通过经天然氨基酸氨基酰化的 O-tRNA 将天然氨基酸并入转录调节剂蛋白质中，产生负性选择标记物的转录。视需要，负性选择标记物包含选择密码子。本发明的正性选择标记物和/或负性选择标记物可包含至少两个选择密码子，其各自或两者可包含至少两个不同的选择密码子或至少两个相同的选择密码子。

转录调节剂蛋白质是（直接地或间接地）与核酸序列（例如反应元件）结合的分子，并且调节操作性地与反应元素连接的序列的转录。转录调节剂蛋白质可为转录活化剂蛋

白质（例如，GAL4、核激素受体、API、CREB、LEF/tcf 家族成员、SMAD、VP16、SP1 等），转录阻遏物蛋白质（例如，核激素受体、Groucho/tle 家族、中锯齿家族（Engrailed family）等），或视环境而定具有两种活性的蛋白质（例如，LEF/tcf、同盒蛋白质（homobox protein）等）。反应元素通常是由转录调节剂蛋白质识别的核酸序列或起与转录调节剂蛋白质一致的作用的其他试剂。

转录调节剂蛋白质的另一个实例是转录活化剂蛋白质，GAL4。参见（例如），A. Laughon 等人，(1984), Identification of two proteins encoded by the *Saccharomyces cerevisiae* GAL4 gene, *Molecular & Cellular Biology* 4:268-275; A. Laughon 和 R. F. Gesteland, (1984), Primary structure of the *Saccharomyces cerevisiae* GAL4 gene, *Molecular & Cellular Biology* 4:260-267; L. Keegan 等人，(1986), Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein, *Science* 231:699-704; 和 M. Ptashne, (1988), How eukaryotic transcriptional activators work, *Nature* 335:683-689。所述 881 氨基酸蛋白质的 N-末端 147 氨基酸形成特异性结合 DNA 序列的 DNA 结合域（DBD）。参见（例如），M. Carey 等人，(1989), An amino-terminal fragment of GAL4 binds DNA as a dimer, *J. Mol. Biol.* 209:423-432; 和 E. Giniger 等人，(1985), Specific DNA binding of GAL4, a positive regulatory protein of yeast, *Cell* 40:767-774。DBD 是通过插入蛋白质序列与当与 DNA 结合时可活化转录的 C-末端 113 氨基酸活化域（AD）连接。参见（例如），J. Ma 和 M. Ptashne, (1987), Deletion analysis of GAL4 defines two transcriptional activating segments, *Cell* 48:847-853; 和 J. Ma 和 M. Ptashne, (1987), The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80, *Cell* 50:137-142。通过将琥珀密码子朝向（例如）含有 GAL4 的 N-末端 DBD 和其 C-末端 AD 的单一多肽的 N-末端 DBD 放置，通过 O-tRNA/O-RS 对的琥珀抑制作用可与通过 GAL4 的转录活化联系。经 GAL4 活化的报告基因可用以执行使用基因的正性和负性选择。

用于负性选择的培养基可包含通过负性选择标记物转化为可检测物质的选择或筛选剂。在本发明的一个方面中，可检测物质是毒性物质。编码负性选择标记物的多核苷酸可为（例如）ura3 基因。举例而言，URA3 报告基因可在含有 GAL4 DNA 结合部位的启动子控制下放置。当负性选择标记物（例如）通过编码具有选择密码子的 GAL4 的多核苷酸的转译而生产时，GAL4 活化 URA3 的转录。负性选择是在包含 5-氟乳清酸（5-FOA）的培养基上完成，所述 5-氟乳清酸通过 ura3 基因的基因产物转化为可检测物质（例如杀死细胞的毒性物质）。参见（例如），J. D. Boeke 等人，(1984), A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoroorotic

acid resistance, *Molecular & General Genetics* 197:345-346); M. Vidal 等人, (1996), Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system., *Proc. Natl. Acad. Sci. U. S. A.* 93:10321-10326; 和 M. Vidal 等人, (1996), Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions., *Proc. Natl. Acad. Sci. U. S. A.* 93:10315-10320.

与正性选择标记物一样, 负性选择标记物也可为各种分子中的任何分子。正性选择标记物和/或负性选择标记物可以是发荧光或在适合的反应物存在下催化发光反应的多肽。举例而言, 负性选择标记物包括(但不限于)(例如)荧光素酶、绿色荧光蛋白质(GFP)、YFP、EGFP、RFP、抗生素抗性基因的产物(例如, 氯霉素乙酰转移酶(CAT))、lacZ 基因的产物、转录调节剂蛋白质等。正性选择标记物和/或负性选择标记物可以通过荧光活化细胞拣选(FACS)或通过发光来检测。正性选择标记物和/或负性选择标记物可以包含基于亲和力的筛选标记物。相同多核苷酸可编码正性选择标记物和负性选择标记物。举例而言, 正性选择步骤、负性选择步骤或正性和负性选择步骤可包括使用报告基因, 其中报告基因是通过荧光活化细胞拣选(FACS)检测。举例而言, 正性选择可先用例如氯霉素乙酰转移酶(CAT)基因的正性选择标记物进行, 其中 CAT 基因包含在 CAT 基因中的例如琥珀终止密码子的选择密码子, 接着进行负性选择筛选, 其是基于在例如 T7 RNA 聚合酶基因的负性标记物中的位置上不能抑制(例如)两个或两个以上的选择密码子来进行。正性选择标记物和负性选择标记物可在例如质粒的相同载体上见到。负性标记物的表达驱动例如绿色荧光蛋白质(GFP)的报告基因的表达。选择和筛选的严格性可以变化, 例如, 使报告基因发荧光所需要的光的强度可以变化。正性选择可用报告基因作为正性选择标记物来进行, 所述报告基因是通过 FACS 筛选, 接着进行负性选择筛选, 其是基于在例如 barnase 基因的负性标记物中的位置上不能抑制(例如)两个或两个以上的选择密码子来进行。

视需要, 报告基因展示在细胞表面上, 例如噬菌体展示(phage display)等等上。例如基于 OmpA 的细胞-表面展示系统的细胞-表面展示依靠例如与大肠杆菌细胞表面上的外膜孔蛋白 OmpA 融合的脊髓灰白质病毒 C3 肽的具体抗原决定基的表达。仅当蛋白质信息中的选择密码子在转译期间被抑制时, 抗原决定基才展示在细胞表面上。然后, 所展示的肽含有由库中的突变体氨酰基-tRNA 合成酶之一识别出的氨基酸, 并且含有对应合成酶基因的细胞可用抵抗含有特定非天然氨基酸的肽而产生的抗体来分离。基于 OmpA 的细胞-表面展示系统是由 Georgiou 等人开发且优化以作为噬菌体展示的替换物。参见, Francisco, J. A., Campbell, R., Iverson, B. L. 和 Georgiou, G. Production and

fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. *Proc. Natl. Acad. Sci. USA* 90:10444-8 (1993).

本发明的其他实施例包括活体外进行一个或一个以上选择步骤。例如合成酶和/或 tRNA 的所选组分可随后被引入细胞中以用于活体内并入非天然氨基酸。

生产 O-RS 和改变合成酶的底物特异性的其他细节可见于标题为“Methods and Compositions for the Production of Orthogonal tRNA-Aminoacyl tRNA Synthetase Pairs”的美国专利申请案 10/126,931 和标题为“EXPANDING THE EUKARYOTIC GENETIC CODE”的 USSN 10/825,867, 其是以引用的方式并入本文中。生产 O-RS 的其他细节可见于 Hamano-Takaku 等人, (2000) A mutant *Escherichia coli* Tyrosyl-tRNA Synthetase Utilizes the Unnatural Amino Acid Azatyrosine More Efficiently than Tyrosine, *Journal of Biological Chemistry*, 275(51):40324-40328; Kiga 等人, (2002), An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system, *PNAS* 99(15): 9715-9723; 和 Francklyn 等人, (2002), Aminoacyl-tRNA synthetases: Versatile players in the changing theater of translation; *RNA*, 8:1363-1372, 各参考文献是以引用的方式并入本文中。

来源和宿主生物体

本发明的转译组分通常由非真核生物体得到。举例而言, 正交 O-tRNA 可由例如古细菌, 诸如詹氏甲烷球菌、横川病毒、诸如富饶盐菌和盐杆菌种 NRC-1 的盐杆菌属、闪烁古生球菌、火球菌、掘越氏热球菌、嗜热菌敏捷气热菌等等; 或真细菌, 诸如大肠杆菌、嗜热细菌、嗜热脂肪芽孢杆菌等等的非真核生物体得到, 而正交 O-RS 可由例如横川病毒、诸如富饶盐菌和盐杆菌种 NRC-1 的盐杆菌属、闪烁古生球菌、火球菌、掘越氏热球菌、嗜热菌敏捷气热菌等等; 或真细菌, 诸如大肠杆菌、嗜热细菌、嗜热脂肪芽孢杆菌等等的非真核生物体得到。在一个实施例中, 也可使用包括(但不限于)植物、藻类、原生生物、真菌、酵母、动物(例如哺乳动物、昆虫、节肢动物等)等等的真核来源。

O-tRNA/O-RS 对的个别组分可由相同生物体或不同生物体得到。在一个实施例中, O-tRNA/O-RS 对是来自相同的生物体。或者, O-tRNA/O-RS 对的 O-tRNA 和 O-RS 是来自不同的生物体。举例而言, O-tRNA 可由(例如)盐杆菌种 *NRC-1* 得到, 并且 O-RS 可由(例如)嗜热自养甲烷杆菌得到。

O-tRNA、O-RS 或 O-tRNA/O-RS 对可经活体内或活体外选择或筛选和/或用于例如

非真核细胞（诸如大肠杆菌细胞）或真核细胞的细胞中以生产具有所选氨基酸（例如非天然氨基酸）的多肽。非真核细胞可来自各种来源，诸如原生种系发生域，包括（但不限于）詹氏甲烷球菌、横川病毒、诸如富饶盐菌和盐杆菌种 *NRC-1* 的盐杆菌属、闪烁古生球菌、火球菌、掘越氏热球菌、嗜热菌敏捷气热菌等等，或可属于真细菌种系发生域，包括（但不限于）大肠杆菌、嗜热细菌、嗜热脂肪芽孢杆菌、荧光极毛杆菌、绿脓杆菌、恶息极毛杆菌等等。真核细胞可来自各种来源，包括（但不限于）植物（例如，诸如单子叶植物或双子叶植物的复杂植物）、藻类、原生生物、真菌、酵母（包括（但不限于）酿酒酵母）；动物（包括（但不限于）哺乳动物、昆虫、节肢动物等）等等。具有本发明的转译组分的细胞的组合物也是本发明的特征。对在一种物种中筛选用于另一种物种的 O-tRNA 和/或 O-RS 来说，又参见，标题为“Expanding the Eukaryotic Genetic Code”的 USSN 10/825,867。

为在宿主细胞中表达具有所选氨基酸的所关注的多肽，所属领域的技术人员可以将编码所关注的多肽的多核苷酸亚克隆到表达载体中，该表达载体含有指导转录的启动子、转录/转译终止子和（如果对于编码蛋白质的核酸来说）用于转译起始的核糖体结合部位的。适合的细菌启动子在所属领域中为熟知的并且描述于（例如）Sambrook 等人和 Ausubel 等人中。

用于表达所关注的多肽的细菌表达系统可在包括（但不限于）大肠杆菌、芽孢杆菌种、荧光极毛杆菌、绿脓杆菌、恶息极毛杆菌和沙门氏菌（沙门氏菌属）中得到（Palva 等人，*Gene* 22:229 - 235 (1983)；Mosbach 等人，*Nature* 302:543 - 545 (1983)）。用于所述表达系统的试剂盒是市售的。用于哺乳动物细胞、酵母和昆虫细胞的真核表达系统在所属领域中是熟知的并且也是市售的。

本发明的 tRNA 和/或 RS 和/或所关注的多肽可以应用于和/或表达于包括（例如）酵母、昆虫细胞、哺乳动物细胞和细菌的许多适合的表达系统中。示范性表达系统的描述提供于下文。

酵母 如本文所使用，术语“酵母”包括能够表达所关注的多肽的各种酵母中的任何酵母。所述酵母包括（但不限于）产子囊孢子酵母（内孢霉目（*Endomycetales*））、担子菌有孢子酵母（*basidiosporogenous yeast*）和属于半知菌（芽生菌（*Blastomycetes*））群组的酵母。产子囊孢子酵母分成 2 个科，即蚀精霉科（*Spermophthoraceae*）和酵母科（*Saccharomycetaceae*）。后者包含 4 个亚科，即裂殖酵母亚科（*Schizosaccharomycoideae*）（例如，裂殖酵母属（genus *Schizosaccharomyces*））、拿逊酵母亚科（*Nadsonioideae*）、*Lipomycoideae* 和 *Saccharomycoideae*（例如毕赤氏酵母属（genera *Pichia*）、克卢费氏酵

母属 (*Kluyveromyces*) 和酵母菌属 (*Saccharomyces*)。担子菌有孢子酵母包括担子菌白冬孢酵母属 (genera *Leucosporidium*)、红冬孢酵母属 (*Rhodosporeidium*)、锁掷酵母属 (*Sporidiobolus*)、线黑粉菌属 (*Filobasidium*) 和香灰拟锁担菌属 (*Filobasidiella*)。属于半知菌 (芽生菌) 群组的酵母分成 2 个科, 即掷孢酵母科 (*Sporobolomycetaceae*) (例如, 掷孢酵母属 (genera *Sporobolomyces*) 和布勒弹孢酵母属 (*Bullera*)) 和隐球酵母科 (*Cryptococcaceae*) (例如念珠菌属 (genus *Candida*))。

供本发明使用的尤其受关注的是以下物种中的物种: 毕赤氏酵母属、克卢费氏酵母属、酵母菌属、裂殖酵母属、汉逊酵母属 (*Hansenula*)、球拟酵母属 (*Torulopsis*) 和念珠菌属, 包括 (但不限于) 巴斯德毕赤酵母 (*P. pastoris*)、*P. guillermondii*、酿酒酵母、嘉士伯酵母 (*S. carlsbergensis*)、糖化酵母 (*S. diastaticus*)、道格拉斯酵母 (*S. douglasii*)、科鲁维尔酵母 (*S. kluyveri*)、诺本斯酵母 (*S. norbensis*)、卵形酵母 (*S. oviformis*)、乳酸克鲁维酵母 (*K. lactis*)、乳糖酶酵母 (*K. fragilis*)、白色念珠菌 (*C. albicans*)、麦芽糖假丝酵母 (*C. maltosa*) 和汉森酵母 (*H. polymorpha*)。酵母通常可从包括 (但不限于) 以下来源的各种来源购得: Yeast Genetic Stock Center、Department of Biophysics and Medical Physics, University of California (Berkeley, CA) 和 American Type Culture Collection (“ATCC”) (Manassas, VA)。

术语“酵母宿主”或“酵母宿主细胞”包括可用作或已经用作用于重组载体或其他转移 DNA 的接受者的酵母。术语包括已经接受重组载体或其他转移 DNA 的原始酵母宿主细胞的子代。应了解, 由于意外的或有意的突变, 单一亲代细胞的子代可以不必在形态学或补充于原始母体的染色体组或总 DNA 方面完全相同。足够相似于有待通过诸如编码所关注的多肽的核苷酸序列存在的相关性质表征的母体的亲代细胞的子代包括在所述定义所指的子代中。

已经开发包括染色体外复制子或合并载体的表达和转形载体, 用于转形到许多酵母宿主中。举例而言, 已经开发用于酿酒酵母的表达载体 (Sikorski 等人, GENETICS (1989) 122:19; Ito 等人, J. BACTERIOL. (1983) 153:163; Hinnen 等人, PROC. NATL. ACAD. SCI. USA (1978) 75:1929); 用于白色念珠菌的表达载体 (Kurtz 等人, MOL. CELL. BIOL. (1986) 6:142); 用于麦芽糖假丝酵母的表达载体 (Kunze 等人, J. BASIC MICROBIOL. (1985) 25:141); 用于汉森酵母的表达载体 (Gleeson 等人, J. GEN. MICROBIOL. (1986) 132:3459; Roggenkamp 等人, MOL. GENETICS AND GENOMICS (1986) 202:302); 用于乳糖酶酵母的表达载体 (Das 等人, J. BACTERIOL. (1984) 158:1165); 用于乳酸克鲁维酵母的表达载体 (De Louvencourt 等人, J. BACTERIOL. (1983) 154:737; Van den Berg 等人,

BIOTECHNOLOGY (NY) (1990) 8:135); 用于 *P. guillermondii* 的表达载体 (Kunze 等人, J. BASIC MICROBIOL. (1985) 25:141); 用于巴斯德毕赤酵母的表达载体 (美国专利第 5,324,639; 4,929,555; 和 4,837,148 号; Cregg 等人, MOL. CELL. BIOL. (1985) 5:3376); 用于粟酒裂殖酵母 (*Schizosaccharomyces pombe*) 的表达载体 (Beach 等人, NATURE (1982) 300:706); 和用于解脂耶罗威亚酵母 (*Y. lipolytica*); 构巢曲霉 (*A. nidulans*) 的表达载体 (Ballance 等人, BIOCHEM. BIOPHYS. RES. COMMUN. (1983) 112:284-89; Tilburn 等人, GENE (1983) 26:205-221; 和 Yelton 等人, PROC. NATL. ACAD. SCI. USA (1984) 81:1470-74); 用于黑霉菌 (*A. niger*) 的表达载体 (Kelly 和 Hynes, EMBO J. (1985) 4:475-479); 用于里氏木霉 (*T. reesia*) 的表达载体 (EP 0 244 234); 和用于诸如脉孢菌属 (*Neurospora*)、青霉属 (*Penicillium*)、弯颈霉属 (*Tolypocladium*) 的丝状真菌的表达载体 (WO 91/00357), 各个参考文献是以引用的方式并入本文中。

酵母载体的控制序列为所属领域的技术人员所知并且包括 (但不限于) 来自诸如以下基因的基因的启动子区域: 醇脱氢酶 (ADH) (EP 0 284 044); 烯醇酶; 葡糖激酶; 葡糖-6-磷酸异构酶; 甘油醛-3-磷酸-脱氢酶 (GAP 或 GAPDH); 己糖激酶; 磷酸果糖激酶; 3-磷酸甘油酸变位酶; 和丙酮酸激酶 (PyK) (EP 0 329 203)。编码酸性磷酸酶的酵母 PHO5 基因也可以提供有效的启动子序列 (Miyano-hara 等人, PROC. NATL. ACAD. SCI. USA (1983) 80:1)。供酵母宿主使用的其他适合的启动子序列可以包括 3-磷酸甘油酸激酶 (Hitzeman 等人, J. BIOL. CHEM. (1980) 255:12073); 和其他糖解酶, 诸如丙酮酸脱羧酶、磷酸丙糖异构酶和磷酸葡糖异构酶 (Holland 等人, Biochemistry (1978) 17:4900; Hess 等人, J. AdvEnzyme Reg. (1969) 7:149) 的启动子。具有通过生长条件来控制的转录的额外优点的诱导型酵母启动子, 可以包括醇脱氢酶 2; 异细胞色素 C; 酸性磷酸酶; 金属硫蛋白; 甘油醛-3-磷酸脱氢酶; 与氮代谢相关的降解酶; 和担负麦芽糖和半乳糖应用的酶的启动子区域。适用于酵母表达的适合的载体和启动子另外描述于 EP 0 073 657 中。

酵母强化子也可以与酵母启动子一起使用。另外, 合成启动子也可以起酵母启动子的作用。举例而言, 酵母启动子的上游活化序列 (UAS) 可以与另一酵母启动子的转录活化区域接合, 产生合成杂交启动子。所述杂交启动子的实例包括与 GAP 转录活化区域连接的 ADH 调节序列。参见美国专利第 4,880,734 和 4,876,197 号, 其是以引用的方式并入本文中。杂交启动子的其他实例所包括的启动子由与诸如 GAP 或 PyK 的糖解酶基因的转录活化区域组合的 ADH2、GAL4、GAL10 或 PHO5 基因的调节序列组成。参见 EP 0 164 556。此外, 酵母启动子可以包括具有结合酵母 RNA 聚合酶并且引发转录的

能力的非酵母源的天然存在的启动子。

可以包含部分酵母表达载体的其他控制元件包括（例如）来自 GAPDH 或烯醇酶基因的终止子（Holland 等人, J. BIOL. CHEM. (1981) 256:1385）。另外, 来自 2μ 质粒源的复制源适用于酵母。适用于酵母的适合的选择基因是存在于酵母质粒中的 *trp1* 基因。参见, Tschumper 等人, GENE (1980) 10:157; Kingsman 等人, GENE (1979) 7:141。 *trp1* 基因提供为没有能力在色氨酸中生长的酵母突变菌株的选择标记物。同样地, 缺 *Leu2* 的酵母菌株 (ATCC 20,622 或 38,626) 是通过带有 *Leu2* 基因的已知质粒补充。

将外源性 DNA 引入酵母宿主中的方法为所属领域的技术人员所知, 并且通常包括（但不限于）用碱性阳离子处理的球形体或完整酵母宿主细胞的转形。举例而言, 酵母的转形可根据 Hsiao 等人, PROC. NATL. ACAD. SCI. USA (1979) 76:3829 和 Van Solingen 等人, J. BACT. (1977) 130:946 中所述的方法进行。然而, 诸如通过核注射、电穿孔或原生质体融合的用于将 DNA 引入细胞中的其他方法, 也可以如 SAMBROOK 等人, MOLECULAR CLONING: A LAB. MANUAL (2001) 中的一般描述使用。然后, 可以使用为所属领域的技术人员所知的标准技术培养酵母宿主细胞。

用于在酵母宿主细胞中表达异源蛋白质的其他方法为所属领域的技术人员所知。通常参见, 美国专利公开案第 20020055169 号、美国专利第 6,361,969; 6,312,923; 6,183,985; 6,083,723; 6,017,731; 5,674,706; 5,629,203; 5,602,034; 和 5,089,398 号; 美国再审专利第 RE37,343 和 RE35,749 号; PCT 公开专利申请案 WO 99/078621; WO 98/37208; 和 WO 98/26080; 欧洲专利申请案 EP 0 946 736; EP 0 732 403; EP 0 480 480; WO 90/10277; EP 0 340 986; EP 0 329 203; EP 0 324 274; 和 EP 0 164 556, 所述专利是以引用的方式并入本文中。又参见 Gellissen 等人, ANTONIE VAN LEEUWENHOEK (1992) 62(1-2):79-93; Romanos 等人, YEAST (1992) 8(6):423-488; Goeddel, METHODS IN ENZYMOLOGY (1990) 185:3-7, 各个参考文献是以引用的方式并入本文中。

酵母宿主菌株在扩增阶段期间, 使用为所属领域的技术人员所知的标准进料分批发酵方法, 在发酵罐中生长。发酵方法可以由于具体的酵母宿主的碳利用路径或表达控制的模式的差异而修改。举例而言, 酵母菌属酵母宿主的发酵可能需要单一葡萄糖进料、复合氮源（例如酪蛋白水解物）和多次维生素增补。相反, 甲基营养型酵母巴斯德毕赤酵母可能需要甘油、甲醇和痕量矿物质进料, 但仅需要简单的铵（氮）盐以达到最佳生长和表达。参见（例如）, 以引用的方式并入本文中的美国专利第 5,324,639 号; Elliott 等人, J. PROTEIN CHEM. (1990) 9:95; 和 Fieschko 等人, BIOTECH. BIOENG. (1987)29:1113。

然而, 所述发酵方法可以具有某些独立于所用酵母宿主菌株的共同特征。举例而言,

通常为碳的生长限制营养物可在扩增阶段期间被添加到发酵罐中以容许最大生长。另外，发酵方法通常使用经设计以含有足够量的碳、氮、基本盐、磷和其他微量养分（维生素、痕量矿物质和盐等）的发酵培养基。适于供毕赤氏酵母属使用的发酵培养基的实例描述于美国专利第 5,324,639 和 5,231,178 号中，所述专利是以引用的方式并入本文中。

经杆状病毒感染的昆虫细胞 术语“昆虫宿主”或“昆虫宿主细胞”指的是可用作或已经用作用于重组载体或其他转移 DNA 的接受者的昆虫。术语包括已经被转染的原始昆虫宿主细胞的子代。应了解，由于意外的或有意的突变，单一亲代细胞的子代可以不必在形态学或补充到原始母体的染色体组或总 DNA 方面完全相同。足够相似于有待通过诸如编码所关注的多肽的核苷酸序列存在的相关性质表征的母体的亲代细胞的子代包括在所述定义所指的子代中。

用于表达所关注的多肽的适合昆虫细胞的选择为所属领域的技术人员所知。若干昆虫物种充分描述于所属领域中并且是市售的，包括埃及伊蚊 (*Aedes aegypti*)、家蚕 (*Bombyx mori*)、果蝇 (*Drosophila melanogaster*)、草地夜蛾 (*Spodoptera frugiperda*) 和粉纹夜蛾 (*Trichoplusia ni*)。在选择用于表达的昆虫宿主时，适合的宿主可以包括展示具有（尤其）良好分泌能力、低蛋白质分解活性和总坚固性的所述宿主。昆虫通常可从包括（但不限于）以下来源的各种来源购得：Insect Genetic Stock Center、Department of Biophysics and Medical Physics, University of California (Berkeley; CA); 和 American Type Culture Collection (“ATCC”) (Manassas, VA)。

通常，经杆状病毒感染的昆虫表达系统的组分包括转移载体，通常为细菌质粒，其含有杆状病毒基因组的片段和用于插入待表达的异源基因的适当的限制性部位；具有与转移载体中的杆状病毒特异性片段同源的序列的野生型杆状病毒（其允许异源基因在杆状病毒基因组中的同源重组）；和适当的昆虫宿主细胞和生长培养基。用于构建载体、转染细胞、挑选溶菌斑、使细胞在培养物中生长诸如此类的物质、方法和技术在所属领域中为已知的并且描述所述技术的手册是可用的。

将异源基因插入转移载体中后，载体和野生型病毒基因组被转染到载体和病毒基因组在其中重组的昆虫宿主细胞中。表达经包装的重组病毒并且鉴别和纯化重组溶菌斑。用于杆状病毒/昆虫细胞表达系统的物质和方法，是来自（例如）Invitrogen Corp. (Carlsbad, CA) 的试剂盒形式市售。所述技术通常为所属领域的技术人员所知并且全部描述于以引用的方式并入本文的 SUMMERS 和 SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN 第 1555 号 (1987) 中。又参见, Richardson, 39 METHODS IN MOLECULAR BIOLOGY: BACULOVIRUS EXPRESSION PROTOCOLS

(1995); AUSUBEL 等人, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY 16.9-16.11 (1994); KING 和 POSSEE, THE BACULOVIRUS SYSTEM: A LABORATORY GUIDE (1992); 和 O'REILLY 等人, BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992)。

实际上,使用杆状病毒/昆虫细胞表达系统的各种异源蛋白质的生产为所属领域的技术人员所知。参见(例如),美国专利第 6,368,825; 6,342,216; 6,338,846; 6,261,805; 6,245,528; 6,225,060; 6,183,987; 6,168,932; 6,126,944; 6,096,304; 6,013,433; 5,965,393; 5,939,285; 5,891,676; 5,871,986; 5,861,279; 5,858,368; 5,843,733; 5,762,939; 5,753,220; 5,605,827; 5,583,023; 5,571,709; 5,516,657; 5,290,686 号; WO 02/06305; WO 01/90390; WO 01/27301; WO 01/05956; WO 00/55345; WO 00/20032; WO 99/51721; WO 99/45130; WO 99/31257; WO 99/10515; WO 99/09193; WO 97/26332; WO 96/29400; WO 96/25496; WO 96/06161; WO 95/20672; WO 93/03173; WO 92/16619; WO 92/02628; WO 92/01801; WO 90/14428; WO 90/10078; WO 90/02566; WO 90/02186; WO 90/01556; WO 89/01038; WO 89/01037; WO 88/07082 号,所述专利以引用的方式并入本文中。

适用于杆状病毒/昆虫细胞表达系统的载体在所属领域中为已知的并且包括(例如)由杆状病毒苜蓿丫纹夜蛾(*Autographa californica*)核型多角体病毒(AcNPV)得到的昆虫表达和转移载体,其为辅助病毒(helper)独立性、病毒表达载体。由所述系统得到的病毒表达载体通常使用强病毒性多角体蛋白基因启动子来驱动异源基因的表达。通常参见, O'Reilly 等人, BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992)。

在将外来基因插入杆状病毒基因组中之前,通常将包含启动子、引子(如果需要)、所关注的编码序列和转录终止序列的上述组分装配成中间错位构筑体(转移载体)。常将中间错位构筑体维持在诸如能够稳定维持在诸如细菌的宿主中的其他染色体元件(例如质粒)的复制子中。复制子将会具有复制系统,因此容许其被维持在用于克隆和扩增的适合的宿主中。更明确地说,质粒可以含有多角体蛋白多聚腺苷酸化信号(Miller, Ann. Rev. Microbiol. (1988) 42:177)和原核抗氨苄青霉素(*amp*)基因和用于在大肠杆菌中选择和繁殖的复制源。

用于将外来基因引入 AcNPV 中的一种常用转移载体是 pAc373。也已经设计了为所属领域的技术人员所知的许多其他载体,包括(例如) pVL985, 其将多角体蛋白起始密码子从 ATG 改变成 ATT, 并且在 ATT 下游的 32 个碱基对处引入 BamHI 克隆部位。参见, Luckow 和 Summers, Virology 170:31 (1989)。其他市售载体包括(例如)

PBlueBac4.5/V5-His; pBlueBacHis2; pMelBac; pBlueBac4.5 (Invitrogen Corp., Carlsbad, CA)。

在插入异源基因后，将转移载体和野生型杆状病毒基因组共转染到昆虫细胞宿主中。用于将异源 DNA 引入杆状病毒病毒中的所要部位中的方法在所属领域中为已知的。参见，SUMMERS 和 SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN 第 1555 号 (1987); Smith 等人, MOL. CELL. BIOL. (1983) 3:2156; Luckow 和 Summers, Virology (1989) 170:31。举例而言，插入可为通过同源复式转线轨道重组，插入到诸如多角体蛋白基因的基因中；插入也可为插入到被工程化成所要杆状病毒基因的限制性内切酶部位中。参见，Miller 等人, Bioessays (1989) 11(4):91。

转染可以通过电穿孔完成。参见 Trotter And Wood, 39 Methods in Molecular BIOLOGY (1995); Mann 和 King, J. Gen. Virol. (1989) 70:3501。或者，脂质体可用以转染具有重组表达载体和杆状病毒的昆虫细胞。参见（例如），Liebman 等人, Biotechniques (1999) 26(1):36; Graves 等人, BIOCHEMISTRY (1998) 37:6050; Nomura 等人, J. BIOL. CHEM. (1998) 273(22): 13570; SCHMIDT 等人, PROTEIN EXPRESSION AND PURIFICATION (1998) 12:323; Siffert 等人, NATURE GENETICS (1998) 18:45; TILKINS 等人, CELL BIOLOGY: A LABORATORY HANDBOOK 145-154 (1998); CAI 等人, PROTEIN EXPRESSION AND PURIFICATION (1997) 10:263; DOLPHIN 等人, NATURE GENETICS (1997) 17:491; Kost 等人, GENE (1997) 190:139; JAKOBSSON 等人, J. BIOL. CHEM. (1996) 271:22203; Rowles 等人, J. BIOL. CHEM. (1996) 271(37):22376; Revere 等人, J. BIOL. CHEM. (1996) 271(39):23607-10; Stanley 等人, J. BIOL. CHEM. (1995) 270:4121; Sisk 等人, J. Virol. (1994) 68(2):766; 和 Peng 等人, BIOTECHNIQUES (1993) 14(2):274。市售脂质体包括（例如）Cellfectin®和 Lipofectin® (Invitrogen, Corp., Carlsbad, CA)。另外，可以使用磷酸钙转染。参见，TROTTER AND WOOD, 39 METHODS IN MOLECULAR BIOLOGY (1995); Kitts, NAR (1990) 18(19):5667; 和，Mann 和 King, J. GEN. VIROL. (1989) 70:3501。

杆状病毒表达载体通常含有杆状病毒启动子。杆状病毒启动子是指任何能够结合杆状病毒 RNA 聚合酶并且引发将编码序列（例如结构基因）下游（3'）转录成 mRNA 的 DNA 序列。启动子将具有通常位于最接近于编码序列的 5'末端的转录起始区域。所述转录起始区域通常包括 RNA 聚合酶结合部位和转录起始部位。杆状病毒启动子也可以具有称为强化子的第二域，如果存在，那么其通常远离结构基因。此外，表达可为调节性或为构成性的。

在感染循环的后期充分转录的结构基因提供尤其有效的启动子序列。实例包括由编码病毒多面体蛋白质的基因(Friesen 等人, *The Regulation of Baculovirus Gene Expression*, THE MOLECULAR BIOLOGY OF BACULOVIRUSES (1986); EP 0 127 839 和 0 155 476) 和编码 p10 蛋白质的基因 (Vlak 等人, *J. Gen. Virol.* (1988) 69:765) 得到的序列。

将新形成的杆状病毒表达载体包装到感染重组杆状病毒中并且随后, 可以通过为所属领域的技术人员所知的技术纯化已生长的溶菌斑。参见, Miller 等人, *Bioessays* (1989) 11(4):91; SUMMERS 和 SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN 第 1555 号(1987)。

已经开发用于感染到若干昆虫细胞中的重组杆状病毒表达载体。举例而言, 已经开发尤其用于埃及伊蚊(ATCC 第 CCL-125 号)、家蚕(ATCC 第 CRL-8910 号)、果蝇(ATCC 第 1963 号)、草地夜蛾和粉蚊夜蛾的重组杆状病毒。参见, Wright, *NATURE* (1986) 321:718; Carbonell 等人, *J. VIROL.* (1985) 56:153; Smith 等人, *MOL. CELL. BIOL.* (1983) 3:2156。通常参见, Fraser 等人, *IN VITRO CELL. DEV. BIOL.* (1989) 25:225。更明确地说, 用于杆状病毒表达载体系统的细胞系通常包括(但不限于) Sf9(草地夜蛾)(ATCC 第 CRL-1711 号)、Sf21(草地夜蛾)(Invitrogen Corp., Cat. 第 11497-013 号(Carlsbad, CA))、Tri-368(粉蚊夜蛾)和 High-Five™ BTI-TN-5B1-4(粉蚊夜蛾)。

用于在杆状病毒/表达中直接表达和融合表达异源多肽的细胞和培养基是市售的, 并且细胞培养技术通常为所属领域的技术人员所知。

大肠杆菌、假单胞菌种和其他原核生物 细菌表达技术为所属领域的技术人员所知。多种载体可得到用于细菌宿主中。载体可以是单拷贝或低或高多拷贝载体。载体可以用于克隆和/或表达。鉴于关于载体、许多载体的工业效用和甚至描述载体和其限制图和特征的手册的丰富的文献, 在此不需要广泛讨论。如所熟知的, 载体通常涉及允许选择的标记物, 所述标记物可以提供细胞毒性剂抗性、原营养或免疫性。常常, 存在提供不同特征的多个标记物。

细菌启动子是指任何能够结合细菌 RNA 聚合酶和引发将编码序列(例如结构基因)下游(3')转录成 mRNA 的 DNA 序列。启动子将具有通常位于最接近于编码序列的 5'末端的转录起始区域。所述转录起始区域通常包括 RNA 聚合酶结合部位和转录起始部位。细菌的启动子也可以具有称为操纵子的第二域, 其可以重叠 RNA 合成开始处的邻近 RNA 聚合酶结合部位。操纵子容许负性调节(诱导型)转录, 因为基因阻遏蛋白可以结合操纵子并且进而抑制特定基因的转录。构成性表达可以在不存在诸如操纵子的负性调节元件时发生。另外, 正性调节可以通过基因活化子蛋白结合序列达成, 如果存在,

那么其通常最接近于 RNA 聚合酶结合序列 (5')。基因活化子蛋白质的实例是代谢活化蛋白 (CAP)，其帮助引发 lac 操纵子在大肠杆菌中的转录[Raibaud 等人, ANNU. REV. GENET. (1984) 18:173]。经调节的表达因此可以是正性的或负性的，进而增强或者减少转录。

编码代谢途径酶的序列提供尤其有效的启动子序列。实例包括由诸如半乳糖、乳糖 (lac) [Chang 等人, Nature (1977) 198:1056]和麦芽糖的糖代谢酶得到的启动子序列。其他实例包括由诸如色氨酸 (trp) 的生物合成酶得到的启动子序列[Goeddel 等人, Nuc. Acids Res. (1980) 8:4057; Yelverton 等人, Nucl. Acids Res. (1981) 9:731; 美国专利第 4,738,921 号; 欧洲专利公开案第 036 776 和 121 775 号, 其是以引用的方式并入本文中]。β-半乳糖苷酶 (bla) 启动子系统[Weissmann (1981) "The cloning of interferon and other mistakes." In Interferon 3 (I. Gresser 编)]、噬菌体 λ PL[Shimatake 等人, Nature (1981) 292:128]和 T5[美国专利第 4,689,406 号]启动子系统也提供有效的启动子序列。诸如 T7 启动子的强启动子可用以诱导高水平的所关注的多肽。所述载体的实例为所属领域的技术人员所知并且包括来自 Novagen 的 pET29 系列和描述于以引用的方式并入本文中的 WO99/05297 中的 pPOP 载体。所述表达系统在宿主中产生高水平的多肽，而不会危害宿主细胞生存性或生长参数。pET19 (Novagen) 是在所属领域中已知的另一种载体。

另外，非天然存在的合成启动子也起细菌启动子的作用。举例而言，一个细菌启动子或噬菌体启动子的转录活化序列可以与另一个细菌启动子或噬菌体启动子的操纵子序列接合，产生合成杂交启动子[美国专利第 4,551,433 号, 所述专利是以引用的方式并入本文中]。举例而言，tac 启动子是通过 lac 阻遏物调节的由 trp 启动子和 lac 操纵子序列组成的杂交 trp-lac 启动子[Amann 等人, GENE (1983) 25:167; de Boer 等人, PROC. NATL. ACAD. SCI. (1983) 80:21]。此外，细菌启动子可包括具有结合细菌 RNA 聚合酶并且引发转录的能力的非细菌性源的天然存在的启动子。非细菌性源的天然存在的启动子也可与相容的 RNA 聚合酶偶合以产生一些基因在原核生物中的高水平表达。细菌磷酸酶 (bacteriophage) T7 RNA 聚合酶/启动子系统是经偶合的启动子系统的实例[Studier 等人, J. Mol. Biol. (1986) 189:113; Tabor 等人, Proc Natl. Acad. Sci. (1985) 82:1074]。另外，杂交启动子也可由噬菌体启动子和大肠杆菌操纵子区域组成(欧洲专利公开案第 267 851 号)。

除功能启动子序列之外，有效的核糖体结合部位也适用于外来基因在原核生物中的表达。在大肠杆菌中，核糖体结合部位称为 Shine-Dalgarno (SD) 序列并且包括起始密码子 (ATG) 和定位于起始密码子的 3-11 个核苷酸上游的长度为 3-9 个核苷酸的序列

[Shine 等人, *Nature* (1975) 254:34]。认为 SD 序列通过 SD 序列与大肠杆菌 16S rRNA 的 3'末端之间的碱基的配对, 而促进 mRNA 与核糖体的结合[Steitz 等人 “Genetic signals and nucleotide sequences in messenger RNA”, In *Biological Regulation and Development: Gene Expression* (Ed. R. F. Goldberger, 1979)]。用弱核糖体结合部位表达真核基因和原核基因[Sambrook 等人 “Expression of cloned genes in *Escherichia coli*”, *Molecular Cloning: A Laboratory Manual*, 1989]。

术语“细菌宿主”或“细菌宿主细胞”指的是可用作或已经用作用于重组载体或其他转移 DNA 的接受者的细菌。术语包括已经转染的原始细菌宿主细胞的子代。应了解, 由于意外的或有意的突变, 单一亲代细胞的子代可以不必在形态学或补充于原始母体的染色体组或总 DNA 方面完全相同。足够相似于有待通过诸如编码 hGH 的核苷酸序列存在的相关性质表征的母体的亲代细胞的子代包括在所述定义所指的子代中。

用于表达多肽的适合寄主细菌的选择为所属领域的技术人员所知。在选择用于表达的细菌宿主时, 适合的宿主可以包括展示具有(尤其)良好内含体形成能力、低蛋白质分解活性和总坚固性的所述宿主。细菌宿主通常可从包括(但不限于)以下来源的各种来源购得: Bacterial Genetic Stock Center、Department of Biophysics and Medical Physics, University of California (Berkeley, CA); 和 American Type Culture Collection (“ATCC”) (Manassas, VA)。工业发酵/医药发酵通常使用由 K 菌株(例如 W3110)得到的细菌或由 B 菌株(例如 BL21)得到的细菌。所述菌株是尤其有效的, 因为其生长参数是十分熟知的和稳固的。另外, 所述菌株是非病原性的, 其在商业上对安全性和环境原因为重要的。适合的大肠杆菌宿主的其他实例包括(但不限于)BL21、DH10B 或其衍生物的菌株。在本发明的方法的另一个实施例中, 大肠杆菌宿主是包括(但不限于)OMP-和 LON-的蛋白酶负菌株。宿主细胞菌株可以是包括(但不限于)荧光极毛杆菌、绿脓杆菌和恶息极毛杆菌的假单胞杆菌种。已知命名为菌株 MB 101 的荧光极毛杆菌生物变种 1 适用于重组生产并且可用于治疗性蛋白质生产工艺。假单胞菌表达系统的实例包括可以宿主菌株(可在 World Wide Web, dow.com上购得的 Midland, MI)购自 Dow Chemical Company 的系统。以引用的方式并入本文中的美国专利第 4,755,465 和 4,859,600 号, 描述假单胞菌株作为用于 hGH 生产的宿主细胞的用途。

一旦重组宿主细胞菌株已经建立(也就是说, 表达构筑体已经被引入宿主细胞中并且具有正确表达构筑体的宿主细胞被分离), 就在适于生产所关注的多肽的条件下培养重组宿主细胞菌株。如所属领域的技术人员显而易见的, 培养重组宿主细胞菌株的方法将视所利用的表达构筑体的性质和宿主细胞的同一性而定。通常使用所属领域中熟知的

方法培养重组宿主菌株。重组宿主细胞通常是在含有碳、氮和无机盐的可吸收源，和（视需要）含有维生素、氨基酸、生长因子和其他为所属领域的技术人员所知的蛋白质培养补充物的液体培养基中培养。用于培养宿主细胞的液体培养基可以视需要含有防止不良微生物生长的抗生素或抗真菌剂和/或包括（但不限于）选择含有表达载体的宿主细胞的抗生素的化合物。

重组宿主细胞可以分批或以连续形式培养，同时细胞收集（在所关注的多肽细胞内积聚的情况下）或培养物上清液的收集以分批或以连续形式进行。为在原核宿主细胞中进行生产，分批培养和细胞收集为优选的。

选择密码子

本发明的选择密码子扩展蛋白质生物合成器的遗传密码子框架。举例而言，选择密码子包括（例如）独特的 3 碱基密码子，无义密码子，诸如包括（但不限于）琥珀密码子（UAG）、赭石密码子或蛋白石密码子（UGA）的终止密码子，非天然密码子，4 碱基（或 4 个以上）密码子，稀有密码子等等。大量选择密码子（例如一个或一个以上、两个或两个以上、3 个或 3 个以上的选择密码子）可被引入所要基因或多核苷酸中。

在一个实施例中，方法涉及选择密码子的用途，该选择密码子为用于活体内并入例如非天然氨基酸的所选氨基酸的终止密码子。举例而言，生产识别终止密码子的 O-tRNA 并且通过 O-RS 用所选氨基酸氨基酰化。所述 O-tRNA 不能通过天然存在的宿主的氨基酰-tRNA 合成酶识别出。惯用定位突变可用以在所关注的多肽中的所关注部位引入终止密码子。参见（例如），Sayers, J.R.等人(1988), *5' - 3' Exonucleases in phosphorothioate - based oligonucleotide - directed mutagenesis. Nucleic Acids Res.* 16:791 - 802。当 O-RS、O-tRNA 和编码所关注的多肽的核酸（例如）活体内组合时，响应终止密码子并入所选氨基酸以得到在规定位置上含有例如非天然氨基酸的所选氨基酸的多肽。在本发明的一个实施例中，用作选择密码子的终止密码子是琥珀密码子 UAG 和/或蛋白石密码子 UGA。举例而言，就识别琥珀密码子的 O-tRNA 的实例来说，参见 SEQ ID NO. : 6，并且就识别蛋白石密码子的 O-tRNA 的实例来说，参见 SEQ ID NO. : 7。UAG 和 UGA 都用作选择密码子的遗传密码可编码 22 种氨基酸，同时保持为最丰富终止信号的赭石无义密码子 UAA。

能在活体内并入例如非天然氨基酸的所选氨基酸而不会显著干扰宿主细胞。举例而言，在诸如大肠杆菌的非真核细胞中，因为 UAG 密码子的抑制效率视例如琥珀抑制基因 tRNA 的 O-tRNA 与释放因子 1 (RF1)（其结合于 UAG 密码子并且引发生长肽从核糖体的释放）之间的竞争而定，所以抑制效率可通过（例如）增加例如抑制基因 tRNA

的 O-tRNA 的表达水平或使用缺乏 RF1 的菌株来调节。在真核细胞中，因为 UAG 密码子的抑制效率视例如琥珀抑制基因 tRNA 的 O-tRNA 与真核释放因子（例如 eRF）（其结合于终止密码子并且引发生长肽从核糖体的释放）之间的竞争而定，所以抑制效率可通过（例如）增加例如抑制基因 tRNA 的 O-tRNA 的表达水平来调节。

非天然氨基酸也能由稀有密码子编码。举例而言，当活体外蛋白质合成反应中的精氨酸浓度降低时，已经证明稀有的精氨酸密码子，即 AGG 对通过由丙氨酸酰化的合成 tRNA 插入 Ala 来说为有效的。参见（例如），Ma 等人，*Biochemistry*, 32:7939 (1993)。在所述情况下，合成 tRNA 与作为次要物种存在于大肠杆菌中的天然存在的 tRNA^{Arg} 竞争。一些生物体不使用所有的三联体密码子。藤黄微球菌（*Micrococcus luteus*）中的未指定的密码子 AGA 已经被应用于在活体外转录/转译提取物中插入氨基酸。参见（例如），Kowal 和 Oliver, *Nucl. Acid. Res.*, 25:4685 (1997)。可产生本发明的组分以在活体内使用所述稀有密码子。

选择密码子也包含（例如）4 个或 4 个以上碱基密码子的扩展密码子，诸如 4 个、5 个、6 个或 6 个以上碱基密码子。4 碱基密码子的实例包括（但不限于）AGGA、CUAG、UAGA、CCCU，诸如此类。5 碱基密码子的实例包括（但不限于）AGGAC、CCCCU、CCCUC、CUAGA、CUACU、UAGGC，诸如此类。特征可包括基于移码抑制使用扩展密码子。4 个或 4 个以上碱基密码子可将（例如）一种或多种包括（但不限于）非天然氨基酸的所选氨基酸插入相同蛋白质中。举例而言，在具有反密码子环，例如具有 CU(X)_n XXXAA 序列（其中 n=1）的例如特殊移码抑制基因 tRNA 的经突变 O-tRNA 的存在下，4 个或 4 个以上碱基密码子读取为单一氨基酸。举例而言，就识别出 4 碱基密码子的 O-tRNA 来说，参见来自 PCT/US04 22061 的 SEQ ID NO. : 6、SEQ ID NO. : 12。在其他实施例中，反密码子环可解码（例如）至少 4-碱基密码子、至少 5-碱基密码子或至少 6-碱基密码子或至少 6 个以上碱基密码子。因为存在 256 种可能的 4-碱基密码子，所以可使用 4 个或 4 个以上碱基密码子在相同细胞中编码多个非天然氨基酸。参见，Anderson 等人，(2002) *Exploring the Limits of Codon and Anticodon Size*, *Chemistry and Biology*, 9:237-244; Magliery, (2001) *Expanding the Genetic Code: Selection of Efficient Suppressors of Four-base Codons and Identification of "Shifty" Four-base Codons with a Library Approach in Escherichia coli*, *J. Mol. Biol.* 307: 755-769。

举例而言，使用活体外生物合成方法，4-碱基密码子已经被用以将非天然氨基酸并入蛋白质中。参见（例如），Ma 等人，(1993) *Biochemistry*, 32:7939; 和 Hohsaka 等人，(1999) *J. Am. Chem. Soc.*, 121:34。CGGG 和 AGGU 用以同时将 2-萘基丙氨酸和赖氨酸的

NBD 衍生物与两种经化学酰化的移码抑制基因 tRNA 一起活体外并入抗生蛋白链菌素中。参见(例如), Hohsaka 等人, (1999) J. Am. Chem. Soc., 121:12194。在活体内研究中, Moore 等人调查了具有 NCUA 反密码子的 tRNA^{Leu} 衍生物抑制 UAGN 密码子(N 可为 U、A、G 或 C)的能力, 并且发现四联体 UAGA 可通过具有 UCUA 反密码子的 tRNA^{Leu}, 以 13 到 26% 的效率解码, 同时在 0 或-1 框中少量解码。参见, Moore 等人, (2000) J. Mol. Biol., 298:195。在一个实施例中, 基于稀有密码子或无义密码子的扩展密码子可用于本发明中, 所述扩展密码子可降低在其他不需要的部位上的错义读取和移码抑制。

对给定系统来说, 选择密码子也可包括天然 3 碱基密码子之一, 其中内源性系统不使用(或很少使用)天然碱基密码子。举例而言, 所述系统包括缺乏识别出天然 3 碱基密码子的 tRNA 的系统, 和/或其中 3 碱基密码子是稀有密码子的系统。

选择密码子视需要包括非天然碱基对。所述非天然碱基对进一步扩展现存的遗传字母表。一个额外的碱基对使三联体密码子的数目从 64 增加到 125。第三碱基对的性质包括稳定的和选择性的碱基配对、以高保真度通过聚合酶有效地酶促并入 DNA 中, 和在初生非天然碱基对的合成后有效的连续的引物延伸。可适宜于方法和组合物的非天然碱基对的描述包括(例如) Hirao 等人, (2002) *An unnatural base pair for incorporating amino acid analogues into protein*, Nature Biotechnology, 20:177 182。又参见, Wu, Y. 等人, (2002) J. Am. Chem. Soc. 124:14626-14630。其他相关公开案列于下文。

对活体内使用来说, 非天然核苷可渗透膜并且经磷酸化以形成对应的三磷酸盐。另外, 增加的遗传信息是稳定的并且不会被细胞酶破坏。Benner 和其他人的早先努力利用不同于典型的 Watson-Crick 对中的所述型式的氢键型式, 其中最值得注意的实例是 iso-C:iso-G 对。参见(例如), Switzer 等人, (1989) J. Am. Chem. Soc., 111:8322; 和 Piccirilli 等人, (1990) Nature, 343:33; Kool, (2000) Curr. Opin. Chem. Biol., 4:602。所述碱基一般说来在某种程度上与天然碱基错配并且不能酶促地复制。Kool 和同事证明, 碱基之间的疏水性包装相互作用能置换氢键以驱动碱基对的形成。参见, Kool, (2000) Curr. Opin. Chem. Biol., 4:602; 和 Guckian 和 Kool, (1998) Angew. Chem. Int. Ed. Engl., 36, 2825。在致力于开发满足所有上述要求的非天然碱基对的过程中, Schultz、Romesberg 和同事系统地合成和研究了一系列非天然疏水性碱基。发现 PICS:PICS 自身对比天然碱基对更稳定, 并且可有效地通过大肠杆菌 DNA 聚合酶 I 的克林诺片段(Klenow fragment, KF)并入 DNA 中。参见(例如), McMinn 等人, (1999) J. Am. Chem. Soc., 121:11585-6; 和 Ogawa 等人, (2000) J. Am. Chem. Soc., 122:3274。3MN:3MN 自身对能通过 KF, 以足够用于生物功能的效率和选择性来合成。参见(例如), Ogawa 等人, (2000) J. Am. Chem.

Soc., 122:8803。然而，两个碱基担当进一步复制的链终止剂。最近，已经开发可用以复制 PICS 自身对的突变 DNA 聚合酶。另外，可复制 7AI 自身对。参见（例如），Tae 等人，(2001) J. Am. Chem. Soc., 123:7439。也已经开发了新颖的金属碱基对，即 Dipic:Py，其在结合铜（II）后形成稳定的对。参见，Meggers 等人，(2000) J. Am. Chem. Soc., 122:10714。因为扩展密码子和非天然密码子本质上与天然密码子正交，所以本发明的方法可以利用所述性质来产生其正交的 tRNA。

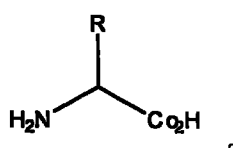
转译旁路系统也可用以将例如非天然氨基酸的所选氨基酸并入所要多肽中。在转译旁路系统中，大的序列被插入基因中但不转译成蛋白质。序列含有充当诱导核糖体跳过序列并且在插入物的下游恢复转译的结构。

所选氨基酸和非天然氨基酸

如本文中所使用，所选氨基酸指的是任何所要求的天然存在的氨基酸或非天然氨基酸。天然存在的氨基酸包括 20 种经遗传编码的 α -氨基酸中的任何一种：丙氨酸、精氨酸、天门冬酰胺、天门冬氨酸、半胱氨酸、谷氨酰胺、谷氨酸、甘氨酸、组氨酸、异白氨酸、白氨酸、赖氨酸、甲硫氨酸、苯丙氨酸、脯氨酸、丝氨酸、苏氨酸、色氨酸、酪氨酸、缬氨酸。在一个实施例中，所选氨基酸是以高保真度，以对给定选择密码子来说大于约 70%的效率、以对给定选择密码子来说大于约 75%的效率、以对给定选择密码子来说大于约 80%的效率、以对给定选择密码子来说大于约 85%的效率、以对给定选择密码子来说大于约 90%的效率、以对给定选择密码子来说大于约 95%的效率，或以对给定选择密码子来说大于约 99%或 99%以上的效率，并入生长多肽链中。

如本文中所使用，非天然氨基酸指的是任何氨基酸、经修饰的氨基酸或不同于硒代半胱氨酸和/或吡咯赖氨酸和以下 20 种经遗传编码的 α -氨基酸的氨基酸类似物：丙氨酸、精氨酸、天门冬酰胺、天门冬氨酸、半胱氨酸、谷氨酰胺、谷氨酸、甘氨酸、组氨酸、异白氨酸、白氨酸、赖氨酸、甲硫氨酸、苯丙氨酸、脯氨酸、丝氨酸、苏氨酸、色氨酸、酪氨酸、缬氨酸。 α -氨基酸的通用结构由式 I 说明：

I



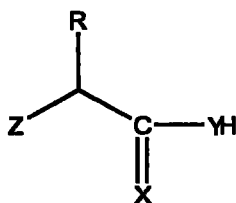
非天然氨基酸通常为具有式 I 的任何结构，其中 R 基团是不同于 20 种天然氨基酸

中所用取代基的任何取代基。就 20 种天然氨基酸的结构来说，参见（例如）L. Stryer 的 *Biochemistry*，第三版，1988, Freeman and Company, New York。注意，本发明的非天然氨基酸可为不同于上述 20 种 α -氨基酸的天然存在的化合物。

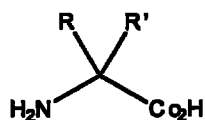
因为本发明的非天然氨基酸通常仅在侧链的结构上不同于天然氨基酸，所以非天然氨基酸以与酰胺键在天然存在的蛋白质中形成的相同方式，与包括（但不限于）天然或非天然的其他氨基酸形成酰胺键。然而，非天然氨基酸具有使其区别于天然氨基酸的侧链基团。举例而言，式 I 中的 R 可以包含烷基-、芳基-、酰基-、酮基-、叠氮基-、羟基-、胍、氰基-、卤基-、酰肼、烯基、炔基、醚、硫醇、硒基-、磺酰基-、硼酸基-、西朋酸基、磷酸基、膦酰基、膦、杂环基、烯酮、亚胺、醛、酯、硫代酸、羟胺、胺，诸如此类或其任何组合。其他非天然存在的氨基酸包括（但不限于）包含可光活化的交联剂的氨基酸、经自旋标记的氨基酸、发荧光的氨基酸、结合金属的氨基酸、含金属的氨基酸、放射性氨基酸、具有新颖的官能团的氨基酸、共价地或非共价地与其他分子相互作用的氨基酸、光致笼罩和/或光致异构化的氨基酸、包含生物素或生物素类似物的氨基酸、诸如经糖取代的丝氨酸的经糖基化的氨基酸、经其他碳水化合物修饰的氨基酸、含酮的氨基酸、包含聚乙二醇或聚醚的氨基酸、经重原子取代的氨基酸、可化学裂解和/或可光致裂解的氨基酸、当与天然氨基酸比较时具有伸长侧链的氨基酸，所述伸长侧链包括（但不限于）聚醚或（包括（但不限于））大于约 5 个或大于约 10 个碳的长链烃、经碳连接的含糖氨基酸、氧化还原活性的氨基酸、含氨基硫代酸的氨基酸和包含一个或一个以上毒性部分的氨基酸。又参见，美国专利申请公开案 2003/0082575 和 2003/0108885，其是以引用的方式并入本文中。非天然氨基酸可以具有用以（例如）将蛋白质连接到固体支撑物的可光活化的交联剂。非天然氨基酸可以具有与氨基酸侧链连接的糖类部分。

除含有新颖侧链的非天然氨基酸之外，（例如）由式 II 和 III 的结构所说明，非天然氨基酸视需要也包含经修饰的骨架结构：

II



III



其中 Z 通常包含 OH、NH₂、SH、NH-R' 或 S-R'；可为相同或不同的 X 和 Y 通常包含 S 或 O，并且视需要为相同或不同的 R 和 R' 通常是选自上文对具有式 I 的非天然氨基酸所描述的 R 基团的成分的同清单以及氢。举例而言，如由式 II 和 III 所说明，非天然氨基酸可以包含在氨基或羧基上的取代。所述类型的非天然氨基酸包括（但不限于） α -羟酸、 α -硫代酸、 α -氨基硫代羧酸酯，例如具有对应于常见 20 种天然氨基酸的侧链或非天然侧链的那些非天然氨基酸。另外，在 α -碳上的取代视需要包括 L-取代、D-取代或 α - α -二取代的氨基酸，诸如 D-谷氨酸、D-丙氨酸、D-甲基-O-酪氨酸、氨基丁酸，诸如此类。其他结构替换物包括诸如脯氨酸类似物以及 3、4、6、7、8 和 9 元环脯氨酸类似物的环状氨基酸，诸如经取代 β -丙氨酸和 γ -氨基丁酸的 β 和 γ 氨基酸。

许多非天然氨基酸是基于诸如酪氨酸、谷氨酰胺、苯丙氨酸，诸如此类的天然氨基酸。酪氨酸类似物包括对位取代的酪氨酸、邻位取代的酪氨酸和间位取代的酪氨酸，其中经取代的酪氨酸包含酮基（包括（但不限于）乙酰基）、苯甲酰基、氨基、胍、羟胺、硫醇基、羧基、异丙基、甲基、C₆-C₂₀ 直链或支链烃、饱和或不饱和烃、O-甲基、聚醚基、硝基等等。另外，也涵盖经多取代的芳基环。谷氨酰胺类似物包括（但不限于） α -羟基衍生物、 γ -取代的衍生物、环状衍生物和经酰胺取代的谷氨酰胺衍生物。苯丙氨酸类似物的实例包括（但不限于）对位取代的苯丙氨酸、邻位取代的苯丙氨酸和间位取代的苯丙氨酸，其中取代基包含羟基、甲氧基、甲基、烯丙基、醛、叠氨基、碘基、溴基、酮基（包括（但不限于）乙酰基）等等。非天然氨基酸的特定实例包括（但不限于）对乙酰基-L-苯丙氨酸、对炔丙基-苯丙氨酸、O-甲基-L-酪氨酸、L-3-(2-萘基)丙氨酸、3-

甲基-苯丙氨酸、O-4-烯丙基-L-酪氨酸、4-丙基-L-酪氨酸、三-O-乙酰基-GlcNAc β -丝氨酸、L-Dopa、氟化苯丙氨酸、异丙基-L-苯丙氨酸、对叠氮基-L-苯丙氨酸、对酰基-L-苯丙氨酸、对苯甲酰基-L-苯丙氨酸、L-磷酸丝氨酸、磷酰基丝氨酸、磷酰基酪氨酸、对碘-苯丙氨酸、对溴苯丙氨酸、对氨基-L-苯丙氨酸、异丙基-L-苯丙氨酸和对炔丙基氧基-苯丙氨酸，诸如此类。各种非天然氨基酸的结构实例提供于（例如）标题为“*In vivo incorporation of unnatural amino acids*”的 WO 2002/085923 中，其是以引用的方式并入本文中。对其他甲硫氨酸类似物来说，又参见，Kiick 等人，(2002) *Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation*, PNAS 99:19-24，其是以引用的方式并入本文中。

在氨基末端并入多肽中的非天然氨基酸可包含为不同于 20 种天然氨基酸中所用取代基中的任何取代基的 R 基团，和不同于通常存在于 α -氨基酸（参见式 I）中的 NH₂ 基团的第二反应性基团。相似的非天然氨基酸可在羧基末端并入具有不同于通常存在于 α -氨基酸（参见式 I）中的 COOH 基团的第二反应性基团。

本发明的非天然氨基酸可以经选择或设计以提供在 20 种天然氨基酸中得不到的其他特征。举例而言，非天然氨基酸可以视需要经设计或选择以改变（例如）所述非天然氨基酸并入其中的蛋白质的生物学性质。举例而言，以下性质可以视需要通过将非天然氨基酸包含于蛋白质中来改变：毒性，体内分解，可溶性，例如热稳定性、水解稳定性、氧化稳定性、抗酶促降解诸如此类的稳定性，纯化和处理的容易性，结构性质，光谱性质，化学的和/或光化性质，催化活性，氧化还原电位，半衰期，与其他分子（例如）共价地或非共价地反应的能力，诸如此类。

各种非天然氨基酸的结构提供于（例如）标题为“*In vivo incorporation of unnatural amino acids*”的 WO 2002/085923 的图 16、17、18、19、26 和 29 中，其是以引用的方式并入本文中。所述实例不意味着以任何方式限制可与本发明的 tRNA 连接的氨基酸。

非天然氨基酸的一个优点是其存在可用以添加其他分子的其他化学部分。所述修饰可在真核或非真核细胞中，活体内或活体外进行。因此，在某些实施例中，后转译修饰是通过非天然氨基酸实现。利用为所属领域的技术人员所知的适于具体反应性基团的化学方法，多肽中的非天然氨基酸可用以将另一分子连接到多肽，包括（但不限于）将包含第二反应性基团的以下物质连接到包含第一反应性基团的至少一种非天然氨基酸：标记、染料、聚合物、水溶性聚合物、聚乙二醇的衍生物、光致交联剂、放射性核素、细胞毒素化合物、药物、亲和标记、光亲和标记、反应性化合物、树脂、第二蛋白质或多肽或多肽类似物、抗体或抗体片段、金属螯合剂、辅助因子、脂肪酸、糖类、多核苷酸、

DNA、RNA、反义多核苷酸、糖类、水溶性树枝状聚合物、环糊精、抑制性核糖核酸、生物材料、纳米粒子、自旋标记物、荧光团、含金属的部分、放射部分、新颖官能团、共价地或非共价地与其他分子相互作用的基团、光致笼罩部分、可光化辐射激发的部分、光致异构化部分、生物素、生物素的衍生物、生物素类似物、并入重原子的部分、可化学裂解的基团、可光致裂解的基团、伸长的侧链、经碳连接的糖、氧化还原性活性剂、氨基硫代酸、毒性部分、经同位素标记的部分、生物物理学探针、发磷光的基团、化学发光基团、电子密基团、磁性基团、插入基团、发色团、能量转移剂、生物活性剂、可检测的标记、小分子、量子点、纳米传导物，或上述物质的任何组合或任何其他所需的化合物或物质。

举例而言，后转译修饰可通过亲核-亲电子反应进行。当前用于选择性修饰蛋白质的大多数反应涉及在亲核与亲电子反应搭配物之间形成共价键，包括（但不限于） α -卤基酮与组氨酸或半胱氨酸侧链的反应。所述情况下的选择性是通过蛋白质中的亲核残基的数量和可达性决定。在本发明的蛋白质中，可使用其他更有选择性的反应，诸如在活体外和活体内，非天然酮-氨基酸与酰肼或氨基氧化化合物的反应。参见（例如），Cornish 等人，(1996) J. Am. Chem. Soc., 118:8150-8151； Mahal 等人，(1997) Science, 276:1125-1128； Wang 等人，(2001) Science 292:498-500； Chin 等人，(2002) J. Am. Chem. Soc. 124:9026-9027； Chin 等人，(2002) Proc. Natl. Acad. Sci., 99:11020-11024； Wang 等人，(2003) Proc. Natl. Acad. Sci., 100:56-61； Zhang 等人，(2003) Biochemistry, 42:6735-6746； 和 Chin 等人，(2003) Science, 301:964-7，其全部以引用的方式并入本文中。所述反应允许用包括荧光团、交联剂、糖类衍生物和细胞毒素分子的大量试剂来选择性标记几乎任何的蛋白质。又参见，标题为“Glycoprotein Synthesis”的美国专利第 6,927,042 号，其是以引用的方式并入本文中。（包括（但不限于））通过叠氮基氨基酸的后转译修饰也能通过 Staudinger 连接（包括（但不限于）用三芳基磷试剂）来进行。参见（例如）Kiick 等人，(2002) *Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation*, PNAS 99:19-24。

非天然氨基酸的化学合成

许多非天然氨基酸可购自（例如）Sigma-Aldrich (St. Louis, MO, USA)、Novabiochem (a division of EMD Biosciences, Darmstadt, Germany) 或 Peptech (Burlington, MA, USA)。不能购得的所述氨基酸视需要如本文中所提供的或使用为所属领域的技术人员所知的标准方法来合成。对有机合成技术来说，参见（例如），Fessenden 和 Fessenden 的 Organic Chemistry (1982, 第二版, Willard Grant Press, Boston Mass.); March 的 Advanced Organic

Chemistry (第三版, 1985, Wiley and Sons, New York); 和 Carey 和 Sundberg 的 Advanced Organic Chemistry (第三版, 部分 A 和 B, 1990, Plenum Press, New York)。描述非天然氨基酸的合成的其他公开案包括 (例如), 标题为 “In vivo incorporation of Unnatural Amino Acids” 的 WO 2002/085923; Matsoukas 等人, (1995) J. Med. Chem., 38, 4660-4669; King, F.E. 和 Kidd, D.A.A. (1949) *A New Synthesis of Glutamine and of γ -Dipeptides of Glutamic Acid from Phthylated Intermediates*. J. Chem. Soc., 3315-3319; Friedman, O.M. 和 Chatterji, R. (1959) *Synthesis of Derivatives of Glutamine as Model Substrates for Anti-Tumor Agents*. J. Am. Chem. Soc. 81, 3750-3752; Craig, J.C. 等人 (1988) *Absolute Configuration of the Enantiomers of 7-Chloro-4 [[4-(diethylamino)-1-methylbutyl]amino]quinoline (Chloroquine)*. J. Org. Chem. 53, 1167-1170; Azoulay, M., Vilmont, M. 和 Frappier, F. (1991) *Glutamine analogues as Potential Antimalarials*, Eur. J. Med. Chem. 26, 201-5; Koskinen, A.M.P. 和 Rapoport, H. (1989) *Synthesis of 4-Substituted Prolines as Conformationally Constrained Ammo Acid Analogues*. J. Org. Chem. 54, 1859-1866; Christie, B.D. 和 Rapoport, H. (1985) *Synthesis of Optically Pure Pipecolates from L-Asparagine. Application to the Total Synthesis of (+)-Apovincamine through Amino Acid Decarboxylation and Iminium Ion Cyclization*. J. Org. Chem. 50:1239-1246; Barton 等人, (1987) *Synthesis of Novel α -Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D- α -Amino-Adipic Acids, L- α -aminopimelic Acid and Appropriate Unsaturated Derivatives*. Tetrahedron 43:4297-4308; 和 Subasinghe 等人, (1992) *Quisqualic acid analogues: synthesis of beta-heterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-sensitized site*. J. Med. Chem. 35:4602-7。又参见, 标题为 “Protein Arrays” 的美国专利公开案第 US 2004/0198637 号, 其是以引用的方式并入本文中。

非天然氨基酸的细胞吸收

非天然氨基酸通过细胞的吸收是通常在设计和选择 (例如) 用于并入蛋白质中的非天然氨基酸时, 通常考虑的一个问题。举例而言, α -氨基酸的高电荷密度暗示所述化合物不可能为可渗透细胞的。天然氨基酸是通过许多以蛋白质为主的运输系统而吸收到细胞中。可进行快速筛选, 其评定哪种非天然氨基酸 (即使有的话) 被细胞吸收。参见 (例如), 在 (例如) 标题为 “Protein Arrays” 的美国专利公开案第 US 2004/0198637 号中的毒性检定, 所述专利是以引用的方式并入本文中; 和 Liu, D.R. 和 Schultz, P.G. (1999) *Progress toward the evolution of an organism with an expanded genetic code*. PNAS United States 96:4780-4785 中的毒性检定。虽然吸收容易用各种检定来分析, 但是设计经受细

胞吸收路径的非天然氨基酸的另一选择为提供活体内产生氨基酸的生物合成路径。

非天然氨基酸的生物合成

许多生物合成路径已经存在于细胞中用于生产氨基酸和其他化合物。虽然具体的非天然氨基酸的生物合成方法可能实质上不存在于（包括（但不限于））细胞中，但是本发明提供所述方法。举例而言，非天然氨基酸的生物合成路径视需要在宿主细胞中，通过添加新的酶或改变现有的宿主细胞路径来产生。其他新的酶视需要为天然存在的酶或人工发展的酶。举例而言，对氨基苯丙氨酸的生物合成（如标题为“*In vivo incorporation of unnatural amino acids*”的 WO 2002/085923 中所呈现）依靠于添加来自其他生物体的已知酶的组合。所述酶的基因能通过用包含基因的质粒使细胞转形而引入细胞中。当在细胞中表达时，基因提供合成所要化合物的酶促路径。视需要添加的酶的类型实例提供于下文的实例中。其他的酶序列见于（例如）基因库中。也视需要以同样方式将人工发展的酶添加到细胞中。如此，操纵细胞器和细胞的资源以生产非天然氨基酸。

各种方法可用于生产供生物合成路径使用或用于发展现有路径的新颖的酶。举例而言，视需要将（例如）如由 Maxygen, Inc.开发的递归性重组（可在 World Wide Web, maxygen.com上得到）用于开发新颖的酶和路径。参见（例如），Stemmer (1994), *Rapid evolution of a protein in vitro by DNA shuffling*, *Nature* 370(4):389-391; 和 Stemmer, (1994), *DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution*, *Proc. Natl. Acad. Sci. USA.*, 91:10747-10751。同样地，视需要将由 Genencor 开发的 DesignPath™（可在 World Wide Web, genencor.com上得到）用于代谢路径工程化，例如用于工程化路径以在细胞中产生 O-甲基-L-酪氨酸。所述技术使用（包括（但不限于））通过功能基因组而鉴别的那些基因的新基因和分子进化和设计的组合，而在宿主生物体中重建现有路径。Diversa Corporation（可在 World Wide Web, diversa.com上得到）也提供快速筛选基因和基因路径的库（包括（但不限于））以产生新的路径的技术。

通常，由本发明的工程化生物合成路径生产的非天然氨基酸，是以包括（但不限于）天然细胞量的足够用于有效的蛋白质生物合成，但不达到影响其他氨基酸的浓度或耗尽细胞资源的程度的浓度来生产。以所述方式在活体内生产的典型浓度是约 10 mM 到约 0.05 mM。一旦细胞由包含用以生产特定路径所需的酶的基因的质粒转形并且产生非天然氨基酸，活体内选择就视需要用以进一步优化用于核糖体蛋白合成和细胞生长的非天然氨基酸的生产。

核酸和多肽序列和变体

如上文和下文所述，本发明提供核酸多核苷酸序列和多肽氨基酸序列（例如 tRNA 和 RS）和（例如）包含所述序列的组合物和方法。所述序列的实例（例如 tRNA 和 RS）揭示于本文中。然而，所属领域的技术人员将了解，本发明不限于揭示本文中（例如实例中）的所述序列。所属领域的技术人员将了解，本发明也提供具有例如编码 O-tRNA 或 O-RS 的本文中描述的功能的许多相关和无关的序列。

本发明提供多肽（O-RS）和例如 O-tRNA 的多核苷酸，编码 O-RS 或其部分的多核苷酸，用以分离氨酰基-tRNA 合成酶无性系的寡核苷酸等。本发明的多核苷酸包括，编码具有一个或一个以上选择密码子的本发明所关注的蛋白质或多肽的多核苷酸。另外，本发明的多核苷酸包括，（例如）包含如 SEQ ED NO. : 4 中所述的核苷酸序列的多核苷酸；与其多核苷酸序列互补的多核苷酸或编码其多核苷酸序列的多核苷酸，或其保守变化。本发明的多核苷酸也包括编码本发明的多肽的多核苷酸。同样地，在高度严格的条件下，在核酸的大体上全部长度上与上文指示的多核苷酸杂交的核酸是本发明的多核苷酸。在一个实施例中，组合物包括本发明的多肽和赋形剂（例如缓冲剂、水、医药学上可接受的赋形剂等）。另外，本发明的多肽包括，（例如）包含如 SEQ ID NO. : 5 中所述的氨基酸序列的多肽；与其多肽序列互补的多肽或编码其多肽序列的多肽，或其保守变化。本发明也提供可与本发明的多肽特异地免疫反应的抗体或抗血清。

在某些实施例中，载体（例如，质粒、粘质粒、噬菌体、细菌、病毒、裸露多核苷酸、结合的多核苷酸等）包含本发明的多核苷酸。在一个实施例中，载体是表达载体。在另一个实施例中，表达载体包括操作性地与本发明的一种或一种以上多核苷酸连接的启动子。在另一个实施例中，细胞包含包括本发明的多核苷酸的载体。

所属领域的技术人员也将了解，所揭示序列的许多变体包括在本发明中。举例而言，产生功能性相同的序列的所揭示序列的保守变化包括在本发明中。变体与至少一种所揭示序列杂交的核酸多核苷酸序列的变体，视为包括在本发明内。如通过（例如）标准序列比较技术所测定的本文中所揭示序列的独特子序列也包括在本发明内。

保守变化

由于遗传密码的简并，“静止的取代”（也就是说，不造成所编码的多肽改变的核酸序列的取代）是编码氨基酸的每种核酸序列的包含特征。同样地，氨基酸序列中的一个或少数氨基酸的“保守氨基酸取代”是经具有高度相似性质的不同氨基酸取代，也易鉴别为高度相似于所揭示的构筑体。各个所揭示序列的所述保守变化是本发明的特征。

具体核酸序列的“保守变化”指的是编码相同或基本上相同的氨基酸序列的核酸，或其中核酸不将氨基酸序列编码成基本上相同的序列的核酸。所属领域的技术人员将认

识到，改变、添加或删除去编码序列中的单一氨基酸或小百分比氨基酸的个别取代、缺失或添加是“经保守修饰的变化”或“经保守修饰的变体”，其中所述改变造成氨基酸的缺失，氨基酸的添加或氨基酸由化学相似的氨基酸取代。因此，本发明的所列多肽序列的“保守变化”包括，用具有相同保守取代基的经保守选择的氨基酸取代多肽序列的小百分比，通常小于5%、更通常小于4%、2%或1%的氨基酸。不改变核酸分子的编码活性的序列的添加，诸如无功能序列的添加，是基本核酸的保守变化。

提供功能性相似的氨基酸的保守取代表为所属领域的技术人员所知。以下八种基团各自含有为用于彼此的保守取代的氨基酸：

- 1) 丙氨酸 (A)、甘氨酸 (G)；
- 2) 天门冬氨酸 (D)、谷氨酸 (E)；
- 3) 天门冬酰胺 (N)、谷氨酰胺 (Q)；
- 4) 精氨酸 (R)、赖氨酸 (K)；
- 5) 异白氨酸 (I)、白氨酸 (L)、甲硫氨酸 (M)、缬氨酸 (V)；
- 6) 苯丙氨酸 (F)、酪氨酸 (Y)、色氨酸 (W)；
- 7) 丝氨酸 (S)、苏氨酸 (T)；和
- 8) 半胱氨酸 (C)、甲硫氨酸 (M)

(参见 (例如), Creighton, *Proteins : Structures and Molecular Properties* (W H Freeman & Co.; 第二版(1993年12月))

核酸杂交

比较性杂交可用以鉴别诸如 SEQ ID NO. : 4 的本发明的核酸，包括本发明的核酸的保守变化，并且所述比较性杂交方法是区别本发明的核酸的优选方法。另外，在高、超高和/或超-超高严格条件下，杂交到由 SEQ ID NO : 4 表示的核酸的目标核酸是本发明的特征。所述核酸的实例包括与给定核酸序列相比，具有一种或少数静止的或保守核酸取代的核酸。

当测试核酸杂交至少 $1/2$ 到探针，同样杂交至少 $1/2$ 到完全匹配的互补目标时，也就是说，在完全匹配的探针以与对不匹配的目标核酸中的任何核酸的杂交所观察到的至少约 5x-10x 高的信噪比而与完全匹配的互补目标结合的条件下，测试核酸以与探针杂交到目标的信噪比的至少 $1/2$ 高的信噪比杂交，被说成特异性地杂交到探针核酸。

核酸当其通常在溶液中缔合时会“杂交”。核酸由于诸如氢键、溶剂排斥、碱基堆积，诸如此类的各种经充分表征的物理化学力而杂交。短语“严格的杂交条件”指的是如所属领域中已知的低离子强度和高温的条件。通常，在严格条件下，探针将杂交到核

酸的复合混合物（包括（但不限于）总体细胞或库的 DNA 或 RNA）中的其目标子序列，但未杂交到复合混合物中的其他序列。核酸杂交的广泛指南可见于 Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* 第 2 章第 I 部分，"Overview of principles of hybridization and the strategy of nucleic acid probe assays,"(Elsevier, New York), 以及见于 Ausubel 等人, *Current Protocols in Molecular Biology* (1995)。英国, 牛津, 牛津大学出版社出版的 Hames and Higgins (1995) Gene Probes 1 IRL (Hames and Higgins 1) 和英国, 牛津, 牛津大学出版社出版的 Hames and Higgins (1995) Gene Probes 2 IRL (Hames and Higgins 2) 提供合成、标记、检测和定量包括寡核苷酸的 DNA 和 RNA 的详细描述。通常, 严格条件选择为比在所定义的离子强度 pH 值下, 特定序列的热力学熔点 (T_m) 低约 5-10°C。 T_m 是 50% 补充到目标的探针与处于平衡的目标序列杂交（因为目标序列在 T_m 下过量存在, 所以 50% 的探针在平衡状态下被占据）的温度（在所定义的离子强度、pH 值和核酸浓度下）。严格条件可以是其中在 pH 7.0 到 8.3 下, 盐浓度小于约 1.0 M 钠离子, 通常约 0.01 到 1.0 M 钠离子浓度（或其他盐）并且对短探针（包括（但不限于）10 到 50 个核苷酸）来说, 温度至少约 30°C 并且对长探针（包括（但不限于）大于 50 个核苷酸）来说, 温度至少约 60°C 的所述严格条件。严格条件也可以由添加诸如甲酰胺的去稳定剂实现。对选择性杂交或特异性杂交来说, 正信号可以是至少 2 次背景杂交, 视需要 10 次背景杂交。示范性严格杂交条件可如下: 50% 甲酰胺、5X SSC 和 1% SDS, 在 42°C 下培养, 或 5X SSC、1% SDS, 在 65°C 下培养, 并在 65°C 下, 在 0.2X SSC 和 0.1% SDS 中洗涤。所述洗涤可执行 5、15、30、60、120 或 120 分钟以上。

在过滤器上, 在南方墨点或北方墨点中, 用于杂交具有超过 100 个互补残基的互补核酸的严格杂交条件的实例是在 42°C 下, 具有 1 mg 的肝素 50% 的福尔马林 (formalin), 以及将杂交执行整夜。严格洗涤条件的实例是在 65°C 下 0.2x SSC 洗涤 15 分钟（参见, Sambrook 等人, *Molecular Cloning, A Laboratory Manual* (第三版 2001) 对 SSC 缓冲液 的描述）。通常, 高严格性的洗涤是在低严格性的洗涤之前进行以除去背景探针信号。实例低严格性的洗涤是在 40°C 下 2x SSC 历时 15 分钟。一般而言, 比在具体的杂交检定中的对无关探针所观察到的信噪比高 5x（或更高）的信噪比指示特异性杂交的检测。

在诸如南方杂交和北方杂交的核酸杂交实验的上下文中的“严格的杂交洗涤条件”是序列依赖性的, 并且在不同的环境参数下而不同。较长序列特别在更高的温度下杂交。核酸的杂交的广泛指南可见于 Tijssen (1993), 同上, 并且见于 Hames 和 Higgins, 1 和 2 中。可容易凭经验测定任何测试核酸的严格杂交和洗涤条件。举例而言, 在测定高严格

的杂交和洗涤条件中，使杂交和洗涤条件逐渐增加（例如，在杂交或洗涤中，通过增加温度、降低盐浓度、增加洗涤剂浓度和/或增加诸如福尔马林的有机溶剂的浓度）直到满足所选的准则组。举例而言，杂交和洗涤条件是逐渐增加，直到探针以与对探针杂交到不匹配目标所观察到的信噪比的至少 5x 高的信噪比与完全匹配的互补目标结合。

选择“极严格的”条件以使其等于具体探针的热熔点 (T_m)。 T_m 是 50% 的测试序列杂交到完全匹配的探针的温度（在所定义的离子强度和 pH 值下）。出于本发明的目的，通常，选择“高严格的”杂交和洗涤条件以使其比在所定义的离子强度和 pH 值下特定序列的 T_m 低约 5°C。

“超高一严格性”杂交和洗涤条件是其中杂交和洗涤条件的严格性增加，直到探针结合到完全匹配的互补目标核酸的信噪比为对杂交到不匹配目标核酸中的任何核酸所观察到的信噪比的至少 10x 高的所述条件。在所述条件下，以完全匹配的互补目标核酸的信噪比的至少 $1/2$ 的信噪比杂交到探针的目标核酸，被说成在超高严格条件下结合到探针。

同样地，甚至更高层次的严格性可通过逐渐增加相关杂交检定的杂交和/或洗涤条件来测定。举例而言，所述条件为，其中杂交和洗涤条件的严格性增加直到探针结合到完全匹配的互补目标核酸的信噪比为对杂交到不匹配目标核酸中的任何核酸所观察到的信噪比的至少 10X、20X、50X、100X 或 500X 或 500X 以上高的条件。在所述条件下，以完全匹配的互补目标核酸的信噪比的 $1/2$ 的信噪比杂交到探针的目标核酸，被说成在超-超高严格条件下结合到探针。

在严格条件下不彼此杂交的核酸，如果其编码的多肽大体上相同的，那么其仍是大体上相同的。这发生在（例如）当使用由遗传密码允许的最大密码子简并来产生核酸的复本时。

独特子序列

在一个方面中，本发明提供包含选自本文中揭示的 O-tRNA 和 O-RS 的序列的核酸中的独特子序列的核酸。与对应于任何已知的 O-tRNA 或 O-RS 核酸序列的核酸相比，独特子序列是独特的。可使用（例如）BLAST 设定缺省参数来执行比对。任何独特子序列适于用作（例如）鉴别本发明的核酸的探针。

同样地，本发明包括多肽，其包含选自本文中揭示的 O-RS 的序列的多肽中的独特子序列。在此，与对应于已知多肽序列中的任何序列的多肽相比，独特子序列是独特的。

本发明也提供目标核酸，其在严格条件下杂交到编码选自 O-RS 的序列的多肽中的独特子序列的独特编码寡核苷酸，其中与对应于任何对照多肽（例如亲本序列，本发明

的合成酶（例如）通过突变自其得到）的多肽相比，独特子序列是独特的。独特序列如上所述测定。

序列比较、同一性和同源性

在两个或两个以上的核酸或多肽序列的情况下，术语“相同”或“同一性”百分比指的是当在比较窗口上或指定区域上比较和比对最大对应性时，按使用下文所述序列比较算法之一（或所属领域的技术人员可用的其他算法）或通过手工比对和目测检查所测量，为相同的或具有规定百分比的相同氨基酸残基或核苷酸的两个或两个以上的序列或子序列。

在两种核酸或多肽（例如，编码 O-tRNA 或 O-RS 的 DNA，或 O-RS 的氨基酸序列）的情况下，短语“大体上相同的”指的是当在比较窗口上或指定区域上比较和比对最大对应性时，如使用序列比较算法（或所属领域的技术人员可用的其他算法）或通过手工比对和目测检查所测量，具有至少约 60%、约 80%、约 90-95%、约 98%、约 99%或 99%以上的核苷酸或氨基酸残基同一性的两个或两个以上的序列或子序列。所述“大体上相同的”序列通常被认为是“同源的”，而与实际祖先无关。“实质同一性”可以在待比较的两个序列的长度为至少约 50 个残基的序列的区域内、至少约 100 个残基的区域内，或至少约 150 个残基的区域内，或全长范围内存在。

对序列比较和同源性测定来说，通常一个序列担当与测试序列比较的参考序列。当使用序列比较算法时，将测试序列和参考序列输入计算机，（必要时）指定子序列坐标并且指定序列算法程序参数。序列比较算法随后基于所指定的程序参数来计算测试序列相对于参考序列的序列同一性百分比。

用于比较的比对序列的方法为所属领域的技术人员所知。用于比较的最佳序列比对法可（例如）通过 Smith 和 Waterman Adv. Appl. Math. 2:482c (1981)的局部同源性算法，通过 Needleman 和 Wunsch J. Mol. Biol. 48:443 (1970)的同源性比对算法，通过 Pearson 和 Lipman Proc. Nat'l. Acad. Sci. USA 85:2444 (1988)的寻找相似性方法的研究，通过所述算法的计算机化实施（Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI 中的 GAP、ESTFIT、FASTA 和 TFASTA）；或通过手工比对和目测检查（参见（例如）Ausubel 等人，*Current Protocols in Molecular Biology* (1995 增刊)）来进行。

适合于测定序列同一性和序列相似性百分比的算法的一个实例是 BLAST 算法和 BLAST 2.0 算法，其分别描述于 Altschul 等人(1997) *Nuc. Acids Res.* 25:3389-3402 中，和 Altschul 等人 J. Mol. Biol. 215:403-410 (1990)中。用于执行 BLAST 分析的软件是在 World

Wide Web, ncbi.nlm.nih.gov 上通过 National Center for Biotechnology Information 而公开获得。所述算法涉及, 首先通过鉴别询问序列中长度为 W 的短字而鉴别高记分序列对 (HSP), 当与数据库序列中具有相同长度的字比对时, 其匹配或满足一些正值阈值记分 T 。 T 称为邻近字记分阈值 (Altschul 等人, 同上)。所述初始邻近字采样数担当启动对含有其的更长 HSP 的搜索的种子。字采样数然后在沿各个序列的两个方向上扩展, 直到累积比对记分可增加为止。核苷酸序列的累积记分使用参数 M (一对匹配残基的回报记分; 始终 >0) 和 N (错配残基的损失记分; 始终 <0) 计算。对氨基酸序列来说, 记分矩阵是用以计算累积分数。当: 累积比对记分从其达到的最大值减少了量 X ; 由于一种或一种以上负性-记分残基比对的积累, 累积记分达到 0 或 0 以下; 或达到各个序列的末端时, 停止字采样数在各个方向上的增加。BLAST 算法参数 W 、 T 和 X 决定比对的敏感性和速度。BLASTN 程序 (对核苷酸序列来说) 使用 11 的字长, 10 的期望值 (E), 100 的截止值, $M = 5$, $N = -4$ 和两个链的比较作为默认值。对氨基酸序列来说, BLASTP 程序使用 3 的字长和 10 的期望值 (E) 作为默认值, 并且 BLOSUM62 记分矩阵 (参见 Henikoff 和 Henikoff (1992) Proc. Natl. Acad. Sci. USA 89:10915) 使用 50 的比对比值 (B), 10 的期望值 (E), $M = 5$, $N = -4$, 和两个链的比较作为默认值。BLAST 算法通常在“低复杂性”过滤器关闭时执行。

除计算序列同一性百分比之外, BLAST 算法也执行两个序列之间的相似性的统计分析 (参见 (例如) Karlin 和 Altschul Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993))。通过 BLAST 算法提供的相似性的一种量度是最小总概率 ($P(N)$), 其指示两个核苷酸或氨基酸序列之间的匹配将偶然发生的概率。举例而言, 如果在测试核酸与参考核酸的比较中的最小总概率小于约 0.2, 小于约 0.01 或小于约 0.001, 那么核酸可以视为与参考序列相似。

突变和其他分子生物学技术

本发明和用于本发明的多核苷酸和多肽可使用分子生物技术来操纵。核苷酸序列可以方便地通过根据惯用方法的定位突变而被修饰。或者, 核苷酸序列可以通过化学合成制备, 包括 (但不限于) 通过使用寡核苷酸合成器制备, 其中寡核苷酸是基于所要多肽的氨基酸序列, 并且优先选择在其中产生重组多肽的宿主细胞中有利的所述密码子而设计。举例而言, 若干编码所要多肽的部分的小寡核苷酸可以通过 PCR、连接或连接链反应合成和装配。参见 (例如), Barany 等人, *Proc. Natl. Acad. Sci.* 88: 189-193 (1991); 美国专利 6,521,427, 其是以引用的方式并入本文中。

本发明利用重组遗传学领域中的常规技术。揭示用于本发明的通用方法的基本文章

包括 Sambrook 等人, *Molecular Cloning, A Laboratory Manual* (第三版 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); 和 *Current Protocols in Molecular Biology* (Ausubel 等人编, 1994))。

描述分子生物技术的一般文章包括 Berger 和 Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology 第 152 卷 Academic Press, Inc., San Diego, CA (Berger); Sambrook 等人, Molecular Cloning-A Laboratory Manual (第二版), 第 1-3 卷, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook")和 Current Protocols in Molecular Biology. F.M. Ausubel 等人编, Current Protocols, a joint venture between Greene Publishing Associates, Inc.和 John Wiley & Sons, Inc., (通过 1999 补充) ("Ausubel")。所述文章描述突变, 载体、启动子的用途和许多其他关于 (例如) 基因或多核苷酸的产生的相关主题, 所述基因或多核苷酸包括用于产生包括所选氨基酸 (例如非天然氨基酸)、正交 tRNA、正交合成酶和其对的蛋白质的选择密码子。

本发明中使用各种类型的突变用于各种目的, 包括 (但不限于) 产生新颖的合成酶或 tRNA, 使 tRNA 分子突变, 产生 tRNA 的库, 使 RS 分子突变, 产生合成酶的库, 产生选择密码子, 将编码所选氨基酸的选择密码子插入所关注的蛋白质或多肽中。所述突变包括 (但不限于) 定位、随机点突变, 同源重组, DNA 滑移或其他递归突变方法, 嵌合建构, 使用含有模板的尿嘧啶的突变、寡核苷酸定位突变, 经硫代磷酸酯修饰的 DNA 突变、使用缺口双链体 DNA 等等的突变, 或其任何组合。其他适合方法包括点错配修复、使用修复缺失宿主菌株的突变, 限制-选择和限制-纯化、缺失突变、通过总基因合成的突变、双链断裂修复, 诸如此类。包括 (但不限于) 涉及嵌合构筑体的突变也包括在本发明中。在一个实施例中, 突变可通过天然存在的分子或经改变或突变的天然存在的分子的包括 (但不限于) 序列, 序列比较, 物理性质, 二级、三级或四级结构, 晶体结构等等已知信息来加以指导。

本文中所见的文章和实例描述所述程序。其他信息见于以下本文中引用的公开案和参考文献中: Ling 等人, *Approaches to DNA mutagenesis: an overview*, Anal Biochem. 254(2): 157-178 (1997); Dale 等人, *Oligonucleotide-directed random mutagenesis using the phosphorothioate method*, Methods Mol. Biol. 57:369-374 (1996); Smith, *In vitro mutagenesis*, Ann. Rev. Genet. 19:423-462 (1985); Botstein 和 Shortle, *Strategies and applications of in vitro mutagenesis*, Science 229:1193-1201 (1985); Carter, *Site-directed mutagenesis*, Biochem. J. 237:1-7 (1986); Kunkel, *The efficiency of oligonucleotide directed mutagenesis*, 在 Nucleic Acids & Molecular Biology (Eckstein, F.和 Lilley, D.M.J.编,

Springer Verlag, Berlin) (1987) 中; Kunkel, *Rapid and efficient site-specific mutagenesis without phenotypic selection*, Proc. Natl. Acad. Sci. USA 82:488-492 (1985); Kunkel 等人, *Rapid and efficient site-specific mutagenesis without phenotypic selection*, Methods in Enzymol. 154, 367-382 (1987); Bass 等人, *Mutant Trp repressors with new DNA-binding specificities*, Science 242:240-245 (1988); Zoller 和 Smith, *Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment*, Nucleic Acids Res. 10:6487-6500 (1982); Zoller 和 Smith, *Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors*, Methods in Enzymol. 100:468-500 (1983); Zoller 和 Smith, *Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template*, Methods in Enzymol. 154:329-350 (1987); Taylor 等人, *The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA*, Nucl. Acids Res. 13: 8749-8764 (1985); Taylor 等人, *The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA*, Nucl. Acids Res. 13: 8765-8785 (1985); Nakamaye 和 Eckstein, *Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis*, Nucl. Acids Res. 14: 9679-9698 (1986); Sayers 等人, *5'-3' Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis*, Nucl. Acids Res. 16:791-802 (1988); Sayers 等人, *Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide*, (1988) Nucl. Acids Res. 16: 803-814; Kramer 等人, *The gapped duplex DNA approach to oligonucleotide-directed mutation construction*, Nucl. Acids Res. 12: 9441-9456 (1984); Kramer 和 Fritz, *Oligonucleotide-directed construction of mutations via gapped duplex DNA*, Methods in Enzymol. 154:350-367 (1987); Kramer 等人, *Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations*, Nucl. Acids Res. 16: 7207 (1988); Fritz 等人, *Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro*, Nucl. Acids Res. 16: 6987-6999 (1988); Kramer 等人, *Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of E. coli*, Cell 38:879-887 (1984); Carter 等人, *Improved oligonucleotide site-directed mutagenesis using M13 vectors*, Nucl. Acids Res. 13: 4431-4443 (1985); Carter, *Improved oligonucleotide-directed mutagenesis using M13 vectors*,

Methods in Enzymol. 154: 382-403 (1987); Eghtedarzadeh 和 Henikoff, *Use of oligonucleotides to generate large deletions*, Nucl. Acids Res. 14: 5115 (1986); Wells 等人, *Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin*, Phil. Trans. R. Soc. Lond. A 317: 415-423 (1986); Nambiar 等人, *Total synthesis and cloning of a gene coding for the ribonuclease S protein*, Science 223: 1299-1301 (1984); Sakmar 和 Khorana, *Total synthesis and expression of a gene for the alpha-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)*, Nucl. Acids Res. 14: 6361-6372 (1988); Wells 等人, *Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites*, Gene 34:315-323 (1985); Grundström 等人, *Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis*, Nucl. Acids Res. 13: 3305-3316 (1985); Mandeck, *Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis*, Proc. Natl. Acad. Sci. USA, 83:7177-7181 (1986); Arnold, *Protein engineering for unusual environments*, Current Opinion in Biotechnology 4:450-455 (1993); Sieber 等人, *Nature Biotechnology*, 19:456-460 (2001); W. P. C. Stemmer, Nature 370, 389-91 (1994); 和 I. A. Lorimer, I. Pastan, Nucleic Acids Res. 23, 3067-8 (1995)。对许多上述方法的其他详细描述可见于 Methods in Enzymology 第 154 卷中, 其也描述对使用各种突变方法故障寻找问题的有效控制。

(例如) 适用于例如使 tRNA 或合成酶的库突变或改变 tRNA 或 RS 的本发明突变的寡核苷酸, 通常是根据由 Beaucage 和 Caruthers, Tetrahedron Letts. 22 (20):1859-1862, (1981)描述的固相亚磷酰胺三酯方法, (例如) 使用如 Needham-VanDevanter 等人, Nucleic Acids Res., 12:6159-6168 (1984)中所述的自动合成器来化学合成。

另外, 基本上任何核酸都可从各种商业来源中的任何来源(诸如 The Midland Certified Reagent Company (mrcr@oligos.com)、The Great American Gene Company (www.genco.com)、ExpressGen Inc. (www.expressgen.com)、Operon Technologies Inc. (Alameda, Calif.))和许多其他来源定制或标准订购。

本发明也涉及真核宿主细胞、非真核宿主细胞和用于通过正交 tRNA/RS 对活体内并入非天然氨基酸的生物体。宿主细胞是由用本发明的多核苷酸, 或包括本发明的多核苷酸(包括(但不限于)可为(例如)克隆载体或表达载体的本发明的载体)的构筑体遗传工程化(包括(但不限于)转形、转导或转染)。举例而言, 正交 tRNA、正交 tRNA 合成酶和待衍生化的蛋白质的编码区域是可操作地与在所宿主细胞中起作用的基因表达控制元件连接。载体可(例如)呈质粒、粘质粒、噬菌体、细菌、病毒、裸露多核

苷酸或结合的多核苷酸形式。载体通过包括以下方法的标准方法引入细胞和/或微生物中：电穿孔（Fromm 等人，*Proc. Natl. Acad. Sci. USA* 82, 5824 (1985))、通过病毒载体进行感染、通过在小珠粒或粒子的基质中或表面上具有核酸的小粒子的高速弹道渗透作用（Klein 等人，*Nature* 327, 70-73 (1987))，和/或诸如此类。

将目标核酸引入细胞中的若干熟知的方法是可用的，任何的方法均可用于本发明中。所述方法包括：将接受者细胞与含有 DNA 的细菌原生质体融合、电穿孔、抛射体轰击和用病毒载体进行感染（下文中进一步讨论）等。细菌细胞可用以扩增含有本发明的 DNA 构筑体的质粒的数量。使细菌生长到对数期并且细菌中的质粒可通过所属领域中已知的各种方法分离（参见（例如）Sambrook）。另外，用于从细菌纯化质粒的试剂盒是市售的（参见（例如）EasyPrep™、FlexiPrep™，都来自 Pharmacia Biotech；StrataClean™，来自 Stratagene；和 QIAprep™，来自 Qiagen）。然后进一步操纵经分离和纯化的质粒来产生用以转染细胞或并入相关载体中以感染生物体的其他质粒。典型的载体含有转录和转译终止子、转录和转译起始序列和用于调节具体的目标核酸的表达的启动子。载体视需要包含基因表达序列盒，所述表达序列盒含有至少一个独立的终止子序列、允许序列盒在真核生物或原核生物或两者（包括（但不限于）穿梭载体）中复制的序列和用于原核系统和真核系统的选择标记物。载体适于在原核生物、真核生物或两者中复制和合并。参见，Gillam 和 Smith, *Gene* 8:81 (1979)；Roberts 等人，*Nature*, 328:731 (1987)；Schneider, E.等人，*Protein Expr. Purif.* 6(1):10-14 (1995)；Ausubel, Sambrook, Berger（所有同上）。用于克隆的细菌和噬菌体的目录（例如）由 ATCC 提供，例如由 ATCC 出版的 *The ATCC Catalogue of Bacteria and Bacteriophage* (1992) Gherna 等人(编)提供。用于定序、克隆及分子生物学的其他方面的其他基本程序和基础的理论观点也见于 Watson 等人(1992) *Recombinant DNA Second Edition Scientific American Books, NY*。另外，基本上任何核酸（和事实上任何经标记的核酸，无论标准或非标准与否）可从各种商业来源中的任何来源（诸如 Midland Certified Reagent Company（可在 World Wide Web, mrcr.com上购得的 Midland, TX）、The Great American Gene Company（可在 World Wide Web, genco.com上购得的 Ramona, CA）、ExpressGen Inc.（可在 World Wide Web, expressgen.com上购得的 Chicago, IL）、Operon Technologies Inc.（Alameda, CA）和许多其他来源定制或标准订购。

经工程化的宿主细胞可在惯用营养培养基中培养，所述营养培养基适当时对于诸如筛选步骤、活化启动子或选择转化体的所述活性而改质。可视需要将所述细胞培养成转基因生物体。其他对（例如）细胞分离和培养（例如，对后续核酸分离）而言有用的参

考文献包括 Freshney (1994) Culture of Animal Cells, a Manual of Basic Technique, 第三版, Wiley-Liss, New York 和其中引用的参考文献; Payne 等人(1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, NY; Gamborg 和 Phillips (编) (1995) Plant Cell, Tissue and Organ Culture: Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York); 和 Atlas 和 Parks (编) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL.

直接将非天然氨基酸活体内并入蛋白质中的能力提供包括(但不限于)以下优点的多种优点: 突变蛋白质的高产率、技术简易性、可能在细胞中或可能在活生物体中研究突变蛋白质, 和在治疗性治疗和诊断用途中使用所述突变蛋白质。使具有各种尺寸、酸性、亲核性、疏水性和其他性质的非天然氨基酸包括在蛋白质中的能力可大大地扩展我们合理地并系统地操纵蛋白质的结构的能力, 以探测蛋白质功能并产生具有新颖性质的新颖蛋白质或生物体。

蛋白质和所关注的多肽

可出于多种目的而进行非天然氨基酸的并入, 包括(但不限于)调节蛋白质结构和/或功能的改变, 改变蛋白酶目标部位的尺寸、酸性、亲核性、氢键、疏水性、可达性, 将部分(包括(但不限于)蛋白质阵列的部分)作为目标, 添加生物活性分子, 连接聚合物, 连接放射性核素, 调节血清半衰期, 调节组织渗透(例如肿瘤), 调节活性传输, 调节组织、细胞或器官特异性或分布, 调节免疫原性, 调节蛋白酶抵抗性等。包括非天然氨基酸的蛋白质可具有增强的或甚至全新的催化或生物物理学性质。举例而言, 以下性质视需要通过将非天然氨基酸包含于蛋白质中而改变: 毒性、体内分解、结构性能、光谱性质、化学和/或光化学性质、催化能力、半衰期(包括(但不限于)血清半衰期)、与其他分子(包括(但不限于))共价地或非共价地反应的能力, 诸如此类。包括包含至少一种非天然氨基酸的蛋白质的组合物适用于(包括(但不限于))新颖治疗、诊断学、催化酶、工业酶、结合蛋白(包括(但不限于)抗体)和(包括(但不限于))蛋白质结构和功能的研究。参见(例如), Dougherty, (2000) *Unnatural Amino Acids as Probes of Protein Structure and Function*, Current Opinion in Chemical Biology, 4:645-652.

蛋白质可具有至少 1 种、至少 2 种、至少 3 种、至少 4 种、至少 5 种、至少 6 种、至少 7 种、至少 8 种、至少 9 种, 或至少 10 种或 10 种以上的非天然氨基酸。非天然氨基酸可为相同的或不同的,(包括(但不限于))在蛋白质中可能存在包含 1 种、2 种、3 种、4 种、5 种、6 种、7 种、8 种、9 种、10 种或 10 种以上的不同的非天然氨基酸的 1 个、2 个、3 个、4 个、5 个、6 个、7 个、8 个、9 个、10 个或 10 个以上的不同部位。

蛋白质可以具有存在于经非天然氨基酸取代的蛋白质中的具体氨基酸的至少一种氨基酸（但少于全部）。对给定的具有 1 种以上的非天然氨基酸的蛋白质来说，非天然氨基酸可为相同的或不同的（包括（但不限于）蛋白质可包括 2 种或 2 种以上不同类型的非天然氨基酸，或可包括两种相同的非天然氨基酸）。对给定的具有 2 种以上的非天然氨基酸的蛋白质来说，非天然氨基酸可为相同的、不同的或为多个具有相同种类的非天然氨基酸与至少 1 种不同的非天然氨基酸的组合。

对在真核细胞中产生所关注的具有至少 1 种非天然氨基酸的蛋白质或多肽来说，蛋白质或多肽通常将包括真核后转译修饰。在某些实施例中，蛋白质包括至少 1 种非天然氨基酸和至少 1 种通过真核细胞在活体内进行的后转译修饰，其中后转译修饰并非通过原核细胞进行。举例而言，后转译修饰包括（包括（但不限于））乙酰化、酰化、脂质修饰、棕榈酰化作用、棕榈酸酯添加、磷酸化、糖脂键修饰、糖基化作用，诸如此类。在另一个方面中，后转译修饰包括前驱物（包括（但不限于）降血钙素前驱物、降血钙素基因相关的肽前驱物、前甲状旁腺激素原激素、前胰岛素原、胰岛素原、前黑素细胞皮质素原、黑素细胞皮质素原，诸如此类）的解脞处理，装配到多子单元蛋白质中或大分子装配，转译到细胞中的另一个部位（包括（但不限于）转译到诸如内质网、高尔基体（Golgi apparatus）、核、溶酶体、过氧化物酶体、线粒体、叶绿体、液泡等等的细胞器中，或通过分泌路径）。在某些实施例中，蛋白质包含分泌序列或定位序列、抗原决定基标签、FLAG 标签、聚组氨酸标签、GST 融合体等等。

在细胞中用所选氨基酸在规定位置生产蛋白质的方法也是本发明的特征。举例而言，方法包括使细胞在适当培养基中生长，其中所述细胞包含至少一个选择密码子并且编码蛋白质的核酸；和提供所选氨基酸；其中细胞另外包含：在细胞中起作用并且识别出选择密码子的正交 tRNA（O-tRNA）；和用所选氨基酸优先氨基酰化 O-tRNA 的正交氨酰基-tRNA 合成酶（O-RS）。通常，O-tRNA 包含在同源合成酶的存在下响应选择密码子的抑制活性。通过所述方法产生的蛋白质也是本发明的特征。

本发明的组合物和通过本发明的方法制得的组合物视需要是在细胞中。本发明的 O-tRNA/O-RS 对或个别组分可随后用于宿主系统的转译机器中，其产生被并入蛋白质中的例如非天然氨基酸的所选氨基酸。标题为“Expanding the Eukaryotic Genetic Code”的专利申请案 USSN 10/825,867 和标题为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS”的 10/126,927 描述了所述方法并且所述专利以引用的方式并入本文。举例而言，当将 O-tRNA/O-RS 对引入例如大肠杆菌的宿主中时，所述对引起所选氨基酸（诸如非天然氨基酸，例如合成氨基酸，诸如白氨酸氨基酸的衍生物）的活体内并入，

所述对可外源地添加到生长培养基中，响应选择密码子而进入蛋白质中。视需要，本发明的组合物可在活体外转译系统中，或在活体内系统中。

包括例如非天然氨基酸的所选氨基酸（和（例如）包括一个或一个以上选择密码子的任何对应的编码核酸）的任何蛋白质（或其部分）可使用本文中的组合物和方法生产。任何多肽适于并入一种或一种以上所选氨基酸。未对鉴别数十万种已知的蛋白质进行尝试，任何蛋白质可经修饰以包括一种或一种以上非天然氨基酸，例如通过调节任何可用的突变方法以在相关转译系统中包括一个或一个以上适当的选择密码子。已知的蛋白质的共用序列储存库包括基因库 EMBL、DDBJ 和 NCBI。其他储存库可容易通过国际互联网搜索而鉴别。

通常，蛋白质是（例如）至少 60%、至少 70%、至少 75%、至少 80%、至少 90%、至少 95%或至少 99%或 99%以上与任何可用的蛋白质（例如，治疗性蛋白质、诊断性蛋白质、工业酶或其部分，诸如此类）相同，并且其包含一种或一种以上所选氨基酸。可经修饰以包含一种或一种以上例如非天然氨基酸的所选氨基酸的治疗性、诊断性和其他蛋白质的实例可见于（但不限于）标题为“Expanding the Eukaryotic Genetic Code”的 USSN 10/825,867 和标题为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS”的美国专利申请案第 10/126,927 号中。

在某些实施例中，本发明的方法和/或组合物中的所关注的蛋白质或多肽（或其部分）是通过核酸编码。通常，核酸包含至少 1 个选择密码子、至少 2 个选择密码子、至少 3 个选择密码子、至少 4 个选择密码子、至少 5 个选择密码子、至少 6 个选择密码子、至少 7 个选择密码子、至少 8 个选择密码子、至少 9 个选择密码子、至少 10 个或 10 个以上的选择密码子。

编码所关注的蛋白质或多肽的基因可使用所属领域的技术人员熟知的的方法和以本文在“突变和其他分子生物学技术”中描述的方法突变，以包括（例如）用于并入例如非天然氨基酸的所选氨基酸的一个或一个以上选择密码子。举例而言，用于所关注的蛋白质的核酸经突变以包括一个或一个以上选择密码子，提供一种或一种以上例如非天然氨基酸的所选氨基酸的插入。本发明包括任何所述变体，例如突变体、例如包括至少一种所选氨基酸的任何蛋白质版本。同样地，本发明也包括对应的核酸，也就是说，具有一个或一个以上编码一种或一种以上所选氨基酸的选择密码子的任何核酸。

为制造包括所选氨基酸的蛋白质，所属领域的技术人员可使用适合于通过正交 tRNA/RS 对活体内并入所选氨基酸的宿主细胞和生物体。宿主细胞是经一种或一种以上表达正交 tRNA、正交 tRNA 合成酶的载体和编码待衍生的蛋白质的载体遗传工程化（例

如, 转形、转导或转染)。所述组分中的各个组分可在同一载体上, 或各个组分可在分离的载体上, 两种组分可在一个载体上并且第三组分可在第二载体上。载体可(例如)呈质粒、粘质粒、噬菌体、细菌、病毒、裸露多核苷酸或结合的多核苷酸形式。

替代系统

已经使用若干策略以在非重组宿主细胞、经突变的宿主细胞或在无细胞系统中, 衍生化具有反应性侧链诸如 Lys、Cys 和 Tyr 的氨基酸会造成赖氨酸向 N²-乙酰基-赖氨酸的转化。化学合成也提供并入非天然氨基酸的直接方法。随着肽片段的酶促连接和天然化学连接的近期发展, 有可能制造较大的蛋白质。参见(例如), P. E. Dawson 和 S. B. H. Kent, Annu. Rev. Biochem., 69:923 (2000)。化学肽连接和天然化学连接描述于美国专利第 6,184,344 号、美国专利公开案第 2004/0138412 号、美国专利公开案第 2003/0208046 号、WO 02/098902 和 WO 03/042235 中, 所述专利是以引用的方式并入本文中。将经所要非天然氨基酸化学酰化的抑制基因 tRNA 添加到能够支撑蛋白质生物合成的活体外提取物的一般活体外生物合成方法, 已经用以将超过 100 种以上的非天然氨基酸部位特异地并入具有任何实际尺寸的各种蛋白质中。参见(例如), V. W. Cornish, D. Mendel 和 P. G. Schultz, Angew. Chem. Int. Ed. Engl., 1995, 34:621 (1995); C.J. Noren, S.J. Anthony-Cahill, M.C. Griffith, P.G. Schultz, *A general method for site-specific incorporation of unnatural amino acids into proteins*, Science 244:182-188 (1989); 和 J.D. Bain, C.G. Glabe, T.A. Dix, A.R. Chamberlin, E.S. Diala, *Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide*, J. Am. Chem. Soc. 111:8013-8014 (1989)。已经将广泛范围的官能团引入蛋白质中以用于研究蛋白质稳定性、蛋白质折叠、酶机制和信号转导。

发展称为选择性压力并入的活体内方法以利用野生型合成酶的杂乱性。参见(例如), N. Budisa, C. Minks, S. Alefelder, W. Wenger, F. M. Dong, L. Moroder 和 R. Huber, FASEB J., 13:41 (1999)。供给细胞具体的天然氨基酸的相关代谢路径被切断的营养缺陷型菌株是在含有有限浓度的天然氨基酸的最小培养基中生长, 而目标基因的转录被抑制。在静止生长期开始时, 天然氨基酸被耗尽并且经非天然氨基酸类似物置换。诱导重组蛋白质的表达造成含有非天然类似物的蛋白质的累积。举例而言, 使用所述策略, 邻位、间位和对位氟苯丙氨酸已经被并入蛋白质中并且在可容易被鉴别的 UV 光谱中显示两个特征肩形突起, 参见(例如), C. Minks, R. Huber, L. Moroder 和 N. Budisa, Anal. Biochem., 284:29 (2000); 三氟甲硫氨酸已经用以置换噬菌体 T4 溶菌酶中的甲硫氨酸以通过 ¹⁹F NMR 研究其与壳寡糖配位体的相互作用, 参见(例如), H. Dewel, E. Daub, V. Robinson 和 J. F. Honek, Biochemistry, 36:3404 (1997); 并且, 三氟白氨酸已被并入代替

白氨酸, 导致白氨酸-拉链蛋白质的热稳定性和化学稳定性增加。参见(例如), Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado 和 D. A. Tirrell, Angew. Chem. Int. Ed. Engl., 40:1494 (2001)。此外, 硒代甲硫氨酸和碲代甲硫氨酸被并入各种重组蛋白质中以在 X 射线结晶学中促进相的溶解。参见(例如), W. A. Hendrickson, J. R. Horton 和 D. M. Lemaster, EMBO J., 9:1665 (1990); J. O. Boles, K. Lewinski, M. Kunkle, J. D. Odom, B. Dunlap, L. Lebioda 和 M. Hatada, Nat. Struct. Biol., 1:283 (1994); N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann 和 R. Huber, Eur. J. Biochem., 230:788 (1995); 和 N. Budisa, W. Karnbrock, S. Steinbacher, A. Humm, L. Prade, T. Neufeind, L. Moroder 和 R. Huber, J. Mol. Biol., 270:616 (1997)。具有烯烃或炔烃官能基的甲硫氨酸类似物也已被有效地并入, 允许通过化学手段额外修饰蛋白质。参见(例如), J. C. van Hest 和 D. A. Tirrell, FEBS Lett., 428:68 (1998); J. C. van Hest, K. L. Kiick 和 D. A. Tirrell, J. Am. Chem. Soc. 122:1282 (2000); 和 K. L. Kiick 和 D. A. Tirrell, Tetrahedron, 56:9487 (2000); 美国专利第 6,586,207 号; 美国专利公开案 2002/0042097, 其是以引用的方式并入本文中。

所述方法的成功取决于通过氨酰基-tRNA 合成酶识别非天然氨基酸类似物, 一般而言, 其需要高的选择性以保证蛋白质转译的保真度。扩展所述方法的范畴的一种方式放宽氨酰基-tRNA 合成酶的底物特异性, 其已经在有限的情况下达成。举例而言, 通过在大肠杆菌苯丙氨酰基-tRNA 合成酶 (PheRS) 中用 Gly 置换 Ala²⁹⁴ 会增加底物结合袋的尺寸, 并且导致通过对氯苯丙氨酸 (p-Cl-Phe) 酰化 tRNAPhe。参见, M. Ibba, P. Kast 和 H. Hennecke, Biochemistry, 33:7107 (1994)。怀有所述突变体 PheRS 的大肠杆菌菌株容许并入对氯苯丙氨酸或对溴苯丙氨酸代替苯丙氨酸。参见(例如), M. Ibba 和 H. Hennecke, FEBS Lett., 364:272 (1995); 和 N. Sharma, R. Furter, P. Kast 和 D. A. Tirrell, FEBS Lett., 467:37 (2000)。同样地, 展示了靠近大肠杆菌酪氨酰 tRNA 合成酶的氨基酸结合部位的点突变 Phe130Ser 容许重氮酪氨酸比酪氨酸更有效地并入。参见, F. Hamano-Takaku, T. Iwama, S. Saito-Yano, K. Takaku, Y. Monden, M. Kitabatake, D. Soil 和 S. Nishimura, J. Biol. Chem., 275:40324 (2000)。

将非天然氨基酸活体内并入蛋白质的另一个策略是修饰具有校对机制的合成酶。所述合成酶不能区别结构上相似于同源天然氨基酸的氨基酸并且因此不能活化所述氨基酸。所述误差在分离部位被校正, 所述分离部位将来自 tRNA 的经错误装填的氨基酸去酰化以维持蛋白质转译的保真度。如果丧失合成酶的校对活性, 那么经错误活化的结构类似物可以避免编辑功能并且被并入。近期已经用缬氨酰基-tRNA 合成酶 (ValRS) 证明了所述方法。参见, V. Doring, H. D. Mootz, L. A. Nangle, T. L. Hendrickson, V. de

Crecy-Lagard, P. Schimmel 和 P. Marliere, Science, 292:501 (2001)。ValRS 可用 Cys、Thr 或氨基丁酸盐 (Abu) 将 tRNA^{Val} 错误氨基酰化; 然后, 通过编辑域使所述非同源氨基酸水解。在大肠杆菌染色体的随机突变后, 选择在 ValRS 的编辑部位具有突变的突变大肠杆菌菌株。所述缺少编辑的 ValRS 用 Cys 不正确装填 tRNA^{Val}。因为 Abu 空间上类似于 Cys (Cys 的 -SH 基团经 Abu 中的 -CH₃ 置换), 所以当所述突变大肠杆菌菌株在 Abu 存在下生长时, 突变 ValRS 也将 Abu 并入蛋白质中。质谱分析展示, 在原生蛋白质中, 约 24% 的缬氨酸在各个缬氨酸位置经 Abu 置换。

固相合成和半合成方法也允许含有新颖氨基酸的大量蛋白质的合成。举例而言, 参见以下公开案和其中引用的参考文献, 所述公开案如下: Crick, F.H.C., Barrett, L. Brenner, S. Watts-Tobin, R. *General nature of the genetic code for proteins*. Nature, 192:1227-1232 (1961); Hofmann, K., Bohn, H. *Studies on polypeptides. XXXVI. The effect of pyrazole-imidazole replacements on the S-protein activating potency of an S-peptide fragment*, J. Am Chem., 88(24):5914-5919 (1966); Kaiser, E.T. *Synthetic approaches to biologically active peptides and proteins including enzymes*, Acc Chem Res, 22:47-54 (1989); Nakatsuka, T., Sasaki, T., Kaiser, E.T. *Peptide segment coupling catalyzed by the semisynthetic enzyme thiosubtilisin*, J Am Chem Soc. 109:3808-3810 (1987); Schnolzer, M., Kent, S B H. *Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease*, Science, 256(5054):221-225 (1992); Chaiken, I.M. *Semisynthetic peptides and proteins*, CRC Crit Rev Biochem. 11(3):255-301 (1981); Offord, R.E. *Protein engineering by chemical means?* Protein Eng., 1(3):151-157 (1987); 和 Jackson, D.Y., Burnier, J., Quan, C., Stanley, M., Tom, J., Wells, J.A. *A Designed Peptide Ligase for Total Synthesis of Ribonuclease A with Unnatural Catalytic Residues*, Science, 266(5183):243 (1994)。

化学修饰已经用以将包括辅助因子、自旋标记物和寡核苷酸的各种非天然侧链活体外引入蛋白质中。参见 (例如), Corey, D.R., Schultz, P.G. *Generation of a hybrid sequence-specific single-stranded deoxyribonuclease*, Science, 238(4832): 1401-1403 (1987); Kaiser, E.T., Lawrence D.S., Rokita, S.E. *The chemical modification of enzymatic specificity*, Annu Rev Biochem, 54:565-595 (1985); Kaiser, E.T., Lawrence, D.S. *Chemical mutation of enzyme active sites*, Science, 226(4674):505-511 (1984); Neet, K.E., Nanci A, Koshland, D.E. *Properties of thiol-subtilisin*, J Biol. Chem., 243(24):6392-6401 (1968); Polgar, L. et M.L. Bender. *A new enzyme containing a synthetically formed active site. Thiol-subtilisin*. J. Am Chem Soc, 88:3153-3154 (1966); 和 Pollack, S.J., Nakayama, G.

Schultz, P.G. *Introduction of nucleophiles and spectroscopic probes into antibody combining sites*, Science, 242(4881):1038-1040 (1988)。

通过免疫反应性定义多肽

因为本发明的多肽提供各种新颖多肽序列（例如，包含在本文中的转译系统中合成的蛋白质的情况下，或在新颖合成酶、标准氨基酸的新颖序列的情况下的所选氨基酸（例如非天然氨基酸）），所以多肽也提供可（例如）在免疫学检定中识别出的新颖结构特征。特异性地结合本发明多肽的抗血清的产生，以及通过所述抗血清结合的多肽都是本发明的特征。如本文中所使用，术语“抗体”包括（但不限于）大体上通过特异性地结合并识别被分析物（抗原）的免疫球蛋白基因或其片段编码的多肽。实例包括多克隆的抗体、单株抗体、嵌合抗体和单链抗体，诸如此类。包括 Fab 片段和由包括噬菌体展示的表达库产生的片段的免疫球蛋白的片段也包括在如本文中所使用的术语“抗体”内。对抗体结构和术语来说，参见（例如），Paul, *Fundamental Immunology*, 第四版，1999, Raven Press, New York。

举例而言，本发明包括 RS 和利用本发明的 tRNA 和/或 RS 制造的蛋白质，其特异性地与抵抗包含氨基酸序列的免疫原而产生的抗体或抗血清结合或特异性地与抵抗包含氨基酸序列的免疫原而产生的抗体或抗血清免疫反应。为消除与其他同源物的交叉反应性，抗体或抗血清是经诸如野生型多肽的可用蛋白质（例如“对照”多肽）减去。在野生型蛋白质对应于核酸时，产生由核酸编码的多肽并且将其用于抗体/抗血清减去用途。

在一个典型形式中，免疫检定使用多克隆抗血清，所述多克隆抗血清经产生以抵抗一种或一种以上多肽或其本质序列（也就是说，至少约 30%的所提供的全长序列）。由蛋白质得到的可能多肽免疫原的组在下文中一起被称为“免疫原性多肽”。所得抗血清视需要经选择以使其对于对照合成酶同源物具有低交叉反应性，并且任何所述交叉反应性是在免疫检定中使用多克隆抗血清之前，通过（例如）免疫吸附用一种或一种以上对照同源物除去。

为生产适用于免疫检定的抗血清，将一种或一种以上免疫原性多肽如本文中所述生产并纯化。举例而言，重组蛋白质可在重组细胞中生产。小鼠的纯系株（因为由于小鼠的实质的遗传同一性，结果为更可重现，所以用于所述检定）是经与诸如弗氏佐剂（Freund' s adjuvant）的标准佐剂组合的免疫原性蛋白质和标准小鼠免疫协议免疫（对抗体产生、免疫检定形式和可用以测定特异免疫反应性的条件的标准描述来说，参见（例如），Harlow 和 Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor

Publications, New York)。

抗体的其他参考文献和论述也见于本文中并且可应用于本文以通过免疫反应性定义多肽。或者，一种或一种以上由本文中揭示的序列得到的合成多肽或重组多肽是与载体蛋白结合并且用作免疫原。

将多克隆血清收集并且在免疫检定中抵抗免疫原性多肽进行滴定，例如，在固相免疫检定中，用一种或一种以上固定在固体支撑物上的免疫原性蛋白质滴定。选择具有 10^6 或 10^6 以上的效价的多克隆抗血清，将其汇集并且用对照合成酶多肽减去以产生经减去汇集滴定的多克隆抗血清。

在比较免疫检定中，测试经减去汇集滴定的多克隆抗血清对于对照同源物的交叉反应性。在所述比较检定中，测定经减去滴定的多克隆抗血清的鉴别结合条件，所述条件产生与结合到对照合成酶同源物相比，高至少约 5-10 倍的将经滴定的多克隆抗血清结合到免疫原性蛋白质的信噪比。换句话说，结合反应的严格性是通过添加诸如白蛋白或脱脂奶粉的非特异竞争剂调整，和/或通过调整盐条件、温度和/或诸如此类来调整。所述结合条件用于后续用于测定测试多肽（待与免疫原性多肽和/或对照多肽相比的多肽）是否通过经汇集减去的多克隆抗血清而特异性结合的检定中。

在另一个实例中，竞争结合形式的免疫检定用于检测测试多肽。举例而言，（如所述）交叉反应抗体是通过与对照多肽的免疫吸附而从所汇集的抗血清混合物中除去。然后将免疫原性多肽固定到暴露于经减去汇集的抗血清的固体支撑物上。将测试蛋白质添加到检定中以竞争与经汇集减去的抗血清的结合。测试蛋白质与固定的蛋白质相比竞争与经汇集减去的抗血清结合的能力，是与被添加到检定中的免疫原性多肽竞争结合（免疫原性多肽有效地与固定免疫原性多肽竞争与经汇集的抗血清的结合）的能力相比较。测试蛋白质的交叉反应性百分比是使用标准算法计算。

在平行检定中，对照蛋白质竞争与经汇集减去的抗血清结合的能力，视需要如与免疫原性多肽竞争与抗血清结合的能力比较来测定。再次，对照多肽的交叉反应性百分比是使用标准算法计算。在与对照多肽比较，测试多肽的交叉反应性百分比高至少 5-10x 时和/或在测试多肽的结合大致在免疫原性多肽的结合的范围内时，测试多肽被说成特异性地结合经汇集减去的抗血清。

一般而言，经免疫吸附和经汇集的抗血清可如本文中所述用于竞争性结合免疫检定中以将任何测试多肽与免疫原性和/或对照多肽进行比较。为进行所述比较，各自在广泛浓度范围下检定免疫原性多肽、测试多肽和对照多肽，并且使用标准技术测定抑制经减去的抗血清与（例如）固定对照蛋白质、测试蛋白质或免疫原性蛋白质的结合的 50%所

需的各种多肽的量。如果在竞争性检定中结合所需的测试多肽的量比所需的免疫原性多肽的量少两倍，那么测试多肽被说成特异地结合对免疫原性蛋白质所产生的抗体，前提是该量比对照多肽的量高至少约 5-10x。

作为特异性的其他测定，经汇集的抗血清视需要全部用免疫原性多肽（而不是对照多肽）免疫吸附，直到很少或没有经所得免疫原性多肽减去汇集的抗血清与用于免疫吸附的免疫原性多肽的结合可检测到。然后测试所述经全部免疫吸附的抗血清与测试多肽的反应性。如果观察到很小或没有反应性（也就是说，对经全部免疫吸附的抗血清与免疫原性多肽的结合来说，观察到不大于 2x 的信噪比），那么测试多肽是通过由免疫原性蛋白质引出的抗血清特异地结合。

对蛋白质、抗体、抗血清等等的其他详细描述可见于标题为“Expanding the Eukaryotic Genetic Code”的 US 10/825,867；标题为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS”的 WO 2002/085923；标题为“Glycoprotein synthesis”的美国专利第 6,927,042 号；和标题为“Protein Arrays”的美国专利公开案第 US 2004/0198637 号，所述专利是以引用的方式并入。

试剂盒

试剂盒也是本发明的特征。举例而言，提供用于在细胞中生产包含至少一种例如非天然氨基酸的所选氨基酸的蛋白质的试剂盒，其中所述试剂盒包括含有编码 O-tRNA 的多核苷酸序列和/或 O-tRNA，和/或编码 O-RS 的多核苷酸序列和/或 O-RS 的容器。在一个实施例中，试剂盒另外至少包括所选氨基酸。在另一个实施例中，试剂盒包括本发明的经氨基酰化的 tRNA。在另一个实施例中，试剂盒另外包含用于生产蛋白质的指导材料。

另一个实例是用于在无细胞转译系统中生产包含至少一种例如非天然氨基酸的所选氨基酸的蛋白质的试剂盒，其中所述试剂盒包括含有编码 O-tRNA 的多核苷酸序列和/或 O-tRNA，和/或编码 O-RS 的多核苷酸序列和/或 O-RS 的容器。在一个实施例中，试剂盒另外包括所选氨基酸。在另一个实施例中，试剂盒包括本发明的经氨基酰化的 tRNA。在另一个实施例中，试剂盒另外包含用于生产蛋白质的指导材料。

实例

提供以下实例以说明，而不是限制所主张的发明。所属领域的技术人员将认识到可在不脱离所主张的发明的范畴的情况下改变的各种非关键参数。

实例 1:

相对于对乙酰基苯丙氨酸的氨酰基-tRNA 合成酶选择

相对于对乙酰基苯丙氨酸（一种非天然编码的氨基酸）筛选两个氨酰基-tRNA 合成酶的 DNA 库。所述库由来自 pBK 质粒中的 *Methanococcus janneschii* 的酪氨酰基 tRNA 合成酶基因的 6 种突变组成。

执行由 5 次选择的交替循环（3 次正性选择、2 次负性选择）组成的选择程序。将库以 1:1 比率组合并且电穿孔到正性选择细胞系（具有正性选择质粒 pREP 的 GeneHog）中并且涂在具有适当的抗生素和非天然编码的氨基酸对乙酰基苯丙氨酸（pAF）的最小培养盘（GMML）上。在 37°C 下将培养盘培养约 40 小时，在所述时间点通过刮削收集细胞。使用 Qiagen Mini-Prep 程序提取 DNA，然后纯化琼脂糖凝胶以分离库质粒 DNA。

然后将所述 DNA 电穿孔到负性选择细胞系（具有负性选择质粒 pBAD 衍生物的 GeneHog）中。将所述转化体涂在具有适当抗生素而不是非天然编码的氨基酸（pAF）LB 盘上。约 17 小时后，通过刮削收集所述细胞并且使用 Qiagen Mini-Prep 程序和琼脂糖凝胶纯化将质粒 DNA 纯化。

利用电穿孔、涂布、收集和 DNA 纯化的相同方法进行选择的后续循环。在选择最后一次（第五次）循环中，用涂在最小培养盘上的经转形的正性选择细胞制成连续稀释液。然后挑取个别群落并且使其在 96 孔块中生长整夜。然后，用变化浓度的具有和不具有非天然氨基酸 pAF 的牛磺胆酸（正性选择抗生素），在最小培养盘上重复涂布所述块。在 37°C 下生长约 40 小时后，目测比较培养盘以确定哪一个群落在最高牛磺胆酸浓度下生长而在不存在非天然编码的氨基酸 pAF 时不生长或不良地生长。使满足所述标准的群落生长整夜。通过 Mini-Prep 和琼脂糖凝胶纯化从培养物中分离 DNA 并且定序。

从对 pAF 的所述选择来看，发现 13 个无性系具有独特的氨基酸序列并且使其经受另外的表征以测定 pAF-tRNA 合成酶的保真度和持续性。

为表征所述合成酶，执行小规模琥珀抑制以展示非天然编码的氨基酸 pAF 被并入多肽中，并且通过 SDS-PAGE 目测结果。挑取单一群落并且使其在 LB 肉汤中生长整夜，然后将其用于接种 50 mL 的 LB。使细胞生长到 OD 为 0.3-0.4，在所述时间点将 1.5 mL 等分试样作为先期诱导点并且将培养物分到 2 个烧瓶中。将 1 mM pAF 添加到一份对分物中并且使两者生长 30 分钟。30 分钟生长后，用 0.2% L-阿拉伯糖诱导 2 个培养物（+/-pAF）并且使其生长 4.5 小时并且记录 OD₆₀₀。然后，取出 +/-pAF 烧瓶的 1.5 mL 等分试样用于 SDS-PAGE 分析。

在 10,000 xg 下将 1.5 mL 等分试样（先期诱导，+pAF、-pAF）离心 10 分钟以使细胞成粒。然后，使细胞悬浮于相对于细胞在收集时的 OD₆₀₀ 的成比例的细菌蛋白质提取试剂（BPER, Pierce）量中。将 DNase I 添加到经溶解的细胞中并且在 4°C 下培养 20 分

钟。然后，将样品与还原剂组合并且加载染料并且在 MES 缓冲液中的 4-12% Bis-TRIS 凝胶上运行 30 分钟。在 DI H₂O 中将凝胶洗涤两次历时 10 分钟并且用考马斯亮蓝 (coomassie blue) 染料染色。比较 +/-pAF 带的 pAF-tRNA RS 引起 pAF 的并入的保真度，并且将+pAF 带与先前所选的 pAF-tRNA RS 比较。

为检查 RS 的持续性，用含有 C-H6 S4am 肌红蛋白 (S4am-Myo) 的质粒执行相同程序。然后，通过 IMAC 纯化 S4am Myo 并且将其送去进行蛋白质定序以测定 pAF 并入量。

在从所述选择识别的 pAF-tRNA RS 中，发现一种合成酶 (E9) 有效地并入 pAF，其以大于 95%效率将 pAF 并入 S4am-Myo 中。并入是通过氨基酸定序测定，而持续性是通过在 SDS-PAGE 凝胶上比较蛋白质带而展示。E9 的核苷酸序列展示于 SEQ ID NO : 4 中，并且 E9 的氨基酸序列展示在 SEQ ID NO : 5 中。

具有与 E9 相似活性的另一突变体被鉴别出，并且具有展示于 SEQ ID NO : 17 中的氨基酸序列。

实例 2:

tRNA 突变

产生 tRNA J17 的 3 种突变体。野生型 J17 的 DNA 序列展示为 SEQ ID NO : 8 并且在美国专利公开案第 2003/0108885 号中展示为 SEQ ID NO : 1 并且在 US 2003/0082575 中展示为 SEQ ID NO : 1 (分别为美国专利申请案第 10/126,931 和 10/126,927 号)，两个专利都以引用的方式并入本文中。如图 1 所示，J17 tRNA 具有在 T ψ C 茎中的 U51 :G63 摆动对。

产生 3 种 J17 突变体 (F12、F13 和 F14) 以在 T ψ C 茎的位置 51 和 63 上产生 Watson-Crick 碱基对。通过重叠 PCR 执行突变，并且将最终构筑体克隆到 pET19 质粒中的 EcoRI 和 NdeI 部位中，所述 pET19 质粒包含编码氨酰基 tRNA 合成酶 E9 的多核苷酸序列 (SEQ ID NO : 4) 和编码具有琥珀密码子取代的人类生长激素 (hGH) 的多核苷酸序列 (SEQ ID NO : 16)。hGH 的表达是受 T7 启动子的控制。

产生用于重叠 PCR 的 2 个片段。第一片段是通过引物延伸获得。用以产生 3 种突变体中的每一种突变体的正向引物的序列是：
GTAACGCTGAATTC CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAAT
CCGCATGGCGC (FTam11; SEQ ID NO: 9)。

为产生 F12 突变体 (51C:63G)，使用以下反向引物：
GATCTGCAGTGGTCCGGCGGGCCGATTTGAACCGGCGCCATGCGGATTTAGAG

TCCGCCGTTCTGC (FTam12; SEQ ID NO: 10)。

为产生 F13 突变体 (51U:63A)，使用以下反向引物：
GATCTGCAGTGGTCCGGCGGGCTGGATTTGAACCAGCGCCATGCGGATTTAGAGT
CCGCCGTTCTGC (FTam13; SEQ ID NO: 11)。

为产生 F14 突变体 (51A:63U)，使用以下反向引物：
GATCTGCAGTGGTCCGGCGGGCAGGATTTGAACCTGCGCCATGCGGATTTAGAGT
CCGCCGTTCTGC (FTam14; SEQ ID NO: 12)。

为产生第二片段，将包含 J17 tRNA 的多核苷酸序列 (SEQ ID NO: 8)、编码 tRNA 合成酶 E9 的多核苷酸序列 (SEQ ID NO: 4) 和编码具有琥珀密码子取代的人类生长激素的多核苷酸序列 (SEQ ID NO: 16) 的质粒 pET19 J17 E9 hGH 用作用于由以下引物组扩增的模板：

CGCCGGACCACTGCAGATCCTTAGCGAAAGCTAAGGATTTTTTTTAAGC (正向引物; FTam15; SEQ ID NO: 13) 和 CAAATTCGTCCATATGGGATTCC (FTam 16; SEQ ID NO: 14)。正向引物用以将序列从 tRNA 的 3'末端扩展到质粒的 Nde I 部位。将所得产物凝胶纯化。

重叠 PCR 的最后步骤涉及正向引物 GTAACGCTGAATTCCCGGCG (FTam17, SEQ ID NO: 15)、反向引物 FTam16 (SEQ ID NO: 14)、第一片段和第二片段。用 EcoR I 和 Nde I 消化经装配的产物并且将其连接到经 EcoR I 和 Nde I 消化的质粒 pET19 J17 E9 hGH 中。各个构筑体的序列是通过定序确认，并且 J17 突变体 tRNA 中的每一个 tRNA 的 DNA 序列展示为 SEQ ID NO: 1 (F12)、SEQ ID NO: 2 (F13) 和 SEQ ID NO: 3 (F14)。tRNA 是以其对应的反向引物命名。

蛋白质表达

编码 tRNA (J17、F12、F13 或 F14) 的质粒各自通过化学方法转形到大肠杆菌菌株 1 和菌株 2 细菌宿主细胞中，并且涂在具有 50 ug/ml 羧苄青霉素的 LB 琼脂培养盘上。在 37°C 下将培养盘培养整夜。对各 tRNA 来说，挑取单一群落以在 37°C 下，在具有 50 ug/ml 羧苄青霉素的 1 ml 2xYT 中开始整夜培养。所述 1 ml 培养物是用以在 37°C 下接种具有 50 ug/ml 羧苄青霉素的两份 10 ml 2xYT 培养物。一份 10 ml 培养物是用 4 mM 对乙酰基苯丙氨酸补充。在 $OD_{600} = 0.7$ 时，用 0.4 mM IPTG 诱导 hGH 表达。在 37°C 下，在 250 rpm 下将细胞培养 4 小时后，通过在 5000 xg 下离心 5 分钟来收集细胞。用经 5 ug/ml DNase I 补充的 B-PER 试剂 (Pierce, Rockford, IL) 溶解细胞。通过 4-12% SDS PAGE 分析总的细胞溶胞产物。

图 2 展示通过 SDS PAGE 对大肠杆菌菌株 1 总细胞溶胞产物的分析。选择密码子在人类生长激素中的抑制作用是使用 J17 或 J17 突变体 (F12、F13、F14) tRNA 和氨酰基 tRNA 合成酶 E9 执行。怀有 J17 突变体的细胞比怀有 J17 的细胞生长更加缓慢。通过 SDS-PAGE, 在不存在 4 mM 对乙酰基苯丙氨酸时, 对 tRNA 突变体未观察到全长 hGH 产物。在 4 mM 对乙酰基苯丙氨酸存在下, 全长产物经 tRNA 突变体中的每一种突变体产生, 证明所述 tRNA 突变体-RS E9 对与大肠杆菌器正交。基于 SDS-PAGE, J17 突变体的经抑制的 hGH 产率比在大肠杆菌菌株 1 中 J17 的经抑制的 hGH 产率高大致 1.5-2 倍。

如图 3 所示, 进一步在大肠杆菌菌株 2 细菌细胞系中测试一种 J17 突变体 F13 的琥珀抑制。在大肠杆菌菌株 2 中, 表达以及琥珀抑制产率相对于在大肠杆菌菌株 1 中的表达以及琥珀抑制产率降低。在不存在对乙酰基苯丙氨酸时, 通过 SDS-PAGE 未观察到全长 hGH 产物。在存在 4 mM 对乙酰基苯丙氨酸时, 对两种 tRNA 都观察到全长 hGH。基于 SDS-PAGE, F13 的经抑制的 hGH 产率比 J17 的经抑制的 hGH 产率高约 3 倍。

进行 J17 与 F13 的比较大的发酵是用大致 1.5 L 的最终体积执行。将编码 J17 tRNA 的质粒和编码 F13 tRNA 的质粒各自转形到大肠杆菌菌株 1 中。各者的最终细胞密度为大致 190 g 湿细胞/升。J17 无性系的 hGH 效价为 347 mg/L 并且 F13 无性系的 hGH 效价为 542 mg/L。

虽然已经出于明确性和理解的目的而相当详细地描述了上述发明, 但是阅读本揭示案, 所属领域的技术人员将显而易见, 可在不脱离本发明真实范畴的情况下, 进行形式和细节的各种变化。举例而言, 所有上述技术和装置都可以各种组合使用。在本申请案中引用的所有公开案、专利、专利申请案和/或其他文献是出于所有目的而以引用的方式全部并入, 该程度就如同已个别指出地各个个别公开案、专利、专利申请案和/或其他文献是出于所有目的而以引用的方式并入一般。

表 1

SEQ ID NO:	标记	序列
1	F12 DNA	CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCA TGGCGCCGGTTCAAATCCGGCCCGCCGGACCA
2	F13 DNA	CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCA TGGCGCTGGTTCAAATCCAGCCCGCCGGACCA
3	F14 DNA	CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCA TGGCGCAGGTTCAAATCCTGCCCGCCGGACCA

4	E9 RS 核酸	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATC AGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAAATC TGCTGTTATAGGTTTTGAACCAAGTGGTAAATAACATTTAGGGCAT TATCTCCAAATAAAAAAGATGATTGATTACAAAATGCTGGATTG ATATAATTATATATTTGGCTGATTACACGCCTATTTAAACCAGAA AGGAGAGTTGGATGAGATTAGAAAAATAGGAGATTATAACAAAA AAGTTTTTGAAGCAATGGGGTTAAAGGCAAATATGTTTATGGAA GTGAACATGGTCTTGATAAGGATTATACACTGAATGTCTATAGATT GGCTTTAAAACTACCTTAAAAAGAGCAAGAAGGAGTATGGA TATAGCAAGAGAGGATGAAAATCCAAAGGTTGCTGAAGTTATCTA TCCAATAATGCAGGTTAATGGATTCAATTATGAGGGCGTTGATGTT GCAGTTGGAGGGATGGAGCAGAGAAAAATACACATGTTAGCAAG GGAGCTTTACCAAAAAAGGTTGTTTGTATCACAAACCCTGTCTTA ACGGGTTTGGATGGAGAAGGAAAGATGAGTTCTTCAAAGGGAAT TTTATAGCTGTTGATGACTCTCCAGAAGAGATTAGGGCTAAGATAA AGAAAGCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCAATAA TGGAGATAGCTAAATACTTCTTGAATATCCTTTAACCATAAAAAG GCCAGAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGA GTTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTTA AAAAATGCTGTAGCTGAAGAACTTATAAAGATTTAGAGCCAATT AGAAAGAGATTATAA
5	E9 RS 氨基酸	MDEFEMKRNTSEIISEEELREVLKKDEKSAVIGFEPGKIHLGHYLQIK KME)LQNAAGFDIIIYLADLHAYLNQKGELE]RKIGDYNKKVFEAMGL KAKYVYGSEHGLDKDYTLNVYRLALKTTLKRARRSMELIAREDENP KVAEVIYPLMQVNGIHYEGVDVAVGGMEQRKIHMLARELLPKKVCCI HNPVLTGLDGEKMSKGNFIAVDDSPPEIRAKIKKAYCPAGVVEG NPMELAKYFLEYPLTIKRPEKFGDLTVNSYEELESFKNKELHPMDL KNAVAEELKILEPIRRL
6	HL(TAG)3 tRNA DNA	CCCAGGGTAGCCAAGCTCGGCCAACGGCGACGGACTCTAAATCCG TTCTCGTAGGAGTTCGAGGGTTCGAATCCCTTCCCTGGGACCA
7	HL(TGA)1 tRNA DNA	GCGGGGGTTGCCGAGCCTGGCCAAAGGCGCCGACTTCAAATCCG GTCCCGTAGGGGTTCCGGGGTCAAATCCCCGCCCGCCACCA
8	J17 詹氏甲烷 球菌 mtRNA ^{Tr} _{CUA} DNA	CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCA TGGCGCTGGTTCAAATCCGGCCCGCCGACCA
9	FTam11 引物	GTAACGCTGAATTCCCGGCGGTAGTTCAGCAGGGCAGAACGGCGG ACTCTAAATCCGCATGGCGC
10	FTam12 引物	GATCTGCAGTGGTCCGGCGGGCCGGATTTGAACCGGCGCCATGCG GATTTAGAGTCCGCCGTTCTGC
11	FTam13 引物	GATCTGCAGTGGTCCGGCGGGCTGGATTTGAACCAGCGCCATGCG GATTTAGAGTCCGCCGTTCTGC
12	FTam14 引物	GATCTGCAGTGGTCCGGCGGGCAGGATTTGAACCTGCGCCATGCG GATTTAGAGTCCGCCGTTCTGC
13	FTam15 引物	AGC
14	FTam16 引物	CAAATTCGTCCATATGGGATTCC
15	FTam17 引物	GTAACGCTGAATTCCCGGCG
16	hGH (DNA)	ATGGGCCACCACCACCACCACCTTCCCAACCATTCCCTTATCCA GGCTTTTTGACAACGCTATGCTCCGCGCCCATCGTCTGCACCAGCT GGCCTTTGACACCTACCAGGAGTTTGAAGAAGCCTAGATCCCAA

		<p>GGAACAGAAGTATTCATTCCTGCAGAACCCCCAGACCTCCCTCTGT TTCTCAGAGTCTATTCCGACACCCTCCAACAGGGAGGAAACACAA CAGAAATCCAACCTAGAGCTGCTCCGCATCTCCCTGCTGCTCATCC AGTCGTGGCTGGAGCCCGTGCAGTTCCTCAGGAGTGTCTTCGCCAA CAGCCTGGTGTACGGCGCCTCTGACAGCAACGTCTATGACCTCCTA AAGGACCTAGAGGAAGGCATCCAAACGCTGATGGGGAGGCTGGA AGATGGCAGCCCCCGGACTGGGCAGATCTTCAAGCAGACCTACAG CAAGTTCGACACAAACTCACACAACGATGACGCACTACTCAAGAA CTACGGGCTGCTCTACTGCTTCAGGAAGGACATGGACAAGGTCGA GACATTCCTGCGCATCGTGCAGTGCCGCTCTGTGGAGGGCAGCTGT GGCTTCTAA</p>
17	E9 的 D286R 突变体	<p>MDEFEMIKRN TSEIISEEEL REVLKKDEKS AVIGFEPGK IHLGHYLQIK KMIDLQNAGF DIIYLADLH AYLNQKGELD EIRKIGDYNK KVFEAMGLKA KYVYGSEHGL DKDYTLNVYR LALKTTLKRA RRSMEIARE DENPKVAEVIYPIMQVNGIH YEGVDVAVGG MEQRKfMMLA RELPKKWCIHNPVLTGLD GEGKMSSSKG NFIAVDDSP EIRAKKKAY CPAGVVEGNP 1MEIAKYFLE YPLTIKRPEK FGGDLTVNSY EELESFLKKN ELHPMRLKNA VAEELKILE PIRKRL</p>

SEQUENCE LISTING

<110> Paulsel, Andrew
 Cho, Ho S.
 <120> 氨基-tRNA合成酶的组合物及其用途
 <130> AMBX-0070.00PCT
 <150> US 60/639,146
 <151> 2004-12-22
 <160> 17
 <170> PatentIn version 3.3
 <210> 1
 <211> 77
 <212> DNA
 <213> 人工序列
 <220>
 <223> 人工序列
 <400> 1
 ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ccggttcaaa 60
 tccgccccgc cggacca 77
 <210> 2
 <211> 77
 <212> DNA
 <213> 人工序列
 <220>
 <223> 人工序列
 <400> 2
 ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ctggttcaaa 60
 tccagccccgc cggacca 77
 <210> 3
 <211> 77
 <212> DNA
 <213> 人工序列
 <220>
 <223> 人工序列
 <400> 3
 ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg cagggtcaaa 60
 tcctgccccgc cggacca 77
 <210> 4
 <211> 921
 <212> DNA
 <213> 人工序列
 <220>
 <223> 人工序列
 <400> 4
 atggacgaat ttgaaatgat aaagagaac acatctgaaa ttatcagcga ggaagagtta 60
 agagagggtt taaaaaaga tgaaaaatct gctgttatag gttttgaacc aagtggtaaa 120
 atacatttag gccattatct ccaataaaa aagatgattg atttacaana tgctggattt 180
 gatataatta tatatttggc tgatttacac gcctatttaa accagaaagg agagtggat 240
 gagattagaa aatagtaga ttataacaaa aaagttttg aagcaatggg gttaaaggca 300
 aatatgttt atggaagtga acatggtctt gataaggatt atacactgaa tgtctataga 360
 ttggctttaa aactacctt aaaaagagca agaaggagta tggaaactat agcaagagag 420
 gatgaaaatc caaaggttgc tgaagttatc tatccaataa tgcaggttaa tgggattcat 480
 tatagggcg ttgatgttgc agttggaggg atggagcaga gaaaaataca catgtagca 540
 agggagcttt taccataaaa gggtgtttgt attcacaacc ctgtctaac gggtttggat 600
 ggagaaggaa agatgagttc ttcaaaaggg aattttatag ctgttgatga ctctccagaa 660
 gagattaggc ctaagataaa gaaagcatac tgcccagctg gagttgttga aggaaatcca 720
 ataattggaga tagctaaata cttcctttaa tacccttaa ccataaaaag gccagaaaaa 780
 tttggtggag atttgacagt taatagctat gaggagttag agagtttatt taaaaataag 840
 gaattgcatc caatggattt aaaaatgct gtagctgaag aactataaa gatttttagag 900
 ccaattagaa agagattata a 921
 <210> 5
 <211> 306
 <212> PRT
 <213> 人工序列
 <220>
 <223> 人工序列
 <400> 5
 Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser
 1 5 10 15
 Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Val
 20 25 30
 Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln
 35 40 45

Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile
 50 55 60
 Tyr Leu Ala Asp Leu His Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp
 65 70 75 80
 Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met
 85 90 95
 Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu His Gly Leu Asp Lys
 100 105 110
 Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys
 115 120 125
 Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro
 130 135 140
 Lys Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Gly Ile His
 145 150 155 160
 Tyr Glu Gly Val Asp Val Ala Val Gly Gly Met Glu Gln Arg Lys Ile
 165 170 175
 His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
 180 185 190
 Asn Pro Val Leu Thr Gly Leu Asp Gly Glu Gly Lys Met Ser Ser Ser
 195 200 205
 Lys Gly Asn Phe Ile Ala Val Asp Asp Ser Pro Glu Glu Ile Arg Ala
 210 215 220
 Lys Ile Lys Lys Ala Tyr Cys Pro Ala Gly Val Val Glu Gly Asn Pro
 225 230 235 240
 Ile Met Glu Ile Ala Lys Tyr Phe Leu Glu Tyr Pro Leu Thr Ile Lys
 245 250 255
 Arg Pro Glu Lys Phe Gly Gly Asp Leu Thr Val Asn Ser Tyr Glu Glu
 260 265 270
 Leu Glu Ser Leu Phe Lys Asn Lys Glu Leu His Pro Met Asp Leu Lys
 275 280 285
 Asn Ala Val Ala Glu Glu Leu Ile Lys Ile Leu Glu Pro Ile Arg Lys
 290 295 300
 Arg Leu
 305

<210> 6
 <211> 88
 <212> DNA
 <213> 盐杆菌种NRC-1
 <400> 6

cccagggtag ccaagctcgg ccaacggcga cggactctaa atccgttctc gtaggagttc 60
 gagggttcga atcccttccc tgggacca 88

<210> 7
 <211> 88
 <212> DNA
 <213> 人工序列
 <220>
 <223> 人工序列
 <400> 7

gcgggggttg ccgagcctgg ccaaaggcgc cggacttcaa atccgggtccc gtaggggttc 60
 cggggttcaa atccccgccc ccgacca 88

<210> 8
 <211> 77
 <212> DNA
 <213> 詹氏甲烷球菌
 <400> 8

ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ctggttcaaa 60
 tccggcccgc cggacca 77

<210> 9
 <211> 65
 <212> DNA
 <213> 人工序列
 <220>
 <223> 人工序列
 <400> 9

gtaacgtga attcccggcg gtagttcagc agggcagaac ggcggactct aaatccgcat 60
 ggcg 65

<210> 10
 <211> 67
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列
 <400> 10
 gatctgcagt ggtccggcgg gccggatttg aaccggcgcc atgcggattt agagtccgcc 60
 gttctgc 67

<210> 11
 <211> 67
 <212> DNA
 <213> 人工序列
 <220>
 <223> 人工序列
 <400> 11
 gatctgcagt ggtccggcgg gctggatttg aaccagcgcc atgcggattt agagtccgcc 60
 gttctgc 67

<210> 12
 <211> 67
 <212> DNA
 <213> 人工序列
 <220>
 <223> 人工序列
 <400> 12
 gatctgcagt ggtccggcgg gcaggatttg aacctgcgcc atgcggattt agagtccgcc 60
 gttctgc 67

<210> 13
 <211> 49
 <212> DNA
 <213> 人工序列
 <220>
 <223> 人工序列
 <400> 13
 cgccggacca ctgcagatcc ttagcgaaaag ctaaggattt tttttaagc 49

<210> 14
 <211> 23
 <212> DNA
 <213> 人工序列
 <220>
 <223> 人工序列
 <400> 14
 caaattcgtc catatgggat tcc 23

<210> 15
 <211> 20
 <212> DNA
 <213> 人工序列
 <220>
 <223> 人工序列
 <400> 15
 gtaacgctga attcccggcg 20

<210> 16
 <211> 600
 <212> DNA
 <213> 智人
 <400> 16
 atgggccacc accaccacca ceacttccca accattecct tatccagget tttgacaac 60
 gctatgtcc ggcgccatcg tctgcaccag ctggccttg acacctacca ggagtttgaa 120
 gaagcctaga tcccaaagga acagaagat tcattcctgc agaaccceca gacctccctc 180
 tgtttctcag agtctattcc gacaccctcc aacagggagg aaacacaaca gaaatccaac 240
 ctagagctgc tcegcatect cctgctgctc atccagtcgt ggctggagcc cgtgcagttc 300
 ctcaggagtg tcttcgcaa cagcctgggtg tacggcgcct ctgacagcaa cgtctatgac 360
 ctctaaag acctagagga aggcattcaa acgctgatgg ggagctgga agatggcagc 420
 ccccgactg ggcagatct caagcagacc tacagcaagt tcgacacaaa ctcacacaac 480
 gatgacgcac tactcaagaa ctacgggctg ctctactgct tcaggaagga catggacaag 540
 gtcgagacat tcctgcgcat cgtgcagtgc cgctctgtgg agggcagctg tggettctaa 600

<210> 17
 <211> 306
 <212> PRT
 <213> 人工序列
 <220>

<223> 人工序列

<400> 17

```

Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser
1      5      10      15
Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Val
20      25      30
Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln
35      40      45
Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile
50      55      60
Tyr Leu Ala Asp Leu His Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp
65      70      75      80
Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met
85      90      95
Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu His Gly Leu Asp Lys
100     105     110
Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys
115     120     125
Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro
130     135     140
Lys Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Gly Ile His
145     150     155     160
Tyr Glu Gly Val Asp Val Ala Val Gly Gly Met Glu Gln Arg Lys Ile
165     170     175
His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
180     185     190
Asn Pro Val Leu Thr Gly Leu Asp Gly Glu Gly Lys Met Ser Ser Ser
195     200     205
Lys Gly Asn Phe Ile Ala Val Asp Asp Ser Pro Glu Glu Ile Arg Ala
210     215     220
Lys Ile Lys Lys Ala Tyr Cys Pro Ala Gly Val Val Glu Gly Asn Pro
225     230     235     240
Ile Met Glu Ile Ala Lys Tyr Phe Leu Glu Tyr Pro Leu Thr Ile Lys
245     250     255
Arg Pro Glu Lys Phe Gly Gly Asp Leu Thr Val Asn Ser Tyr Glu Glu
260     265     270
Leu Glu Ser Leu Phe Lys Asn Lys Glu Leu His Pro Met Arg Leu Lys
275     280     285
Asn Ala Val Ala Glu Glu Leu Ile Lys Ile Leu Glu Pro Ile Arg Lys
290     295     300
Arg Leu
305

```

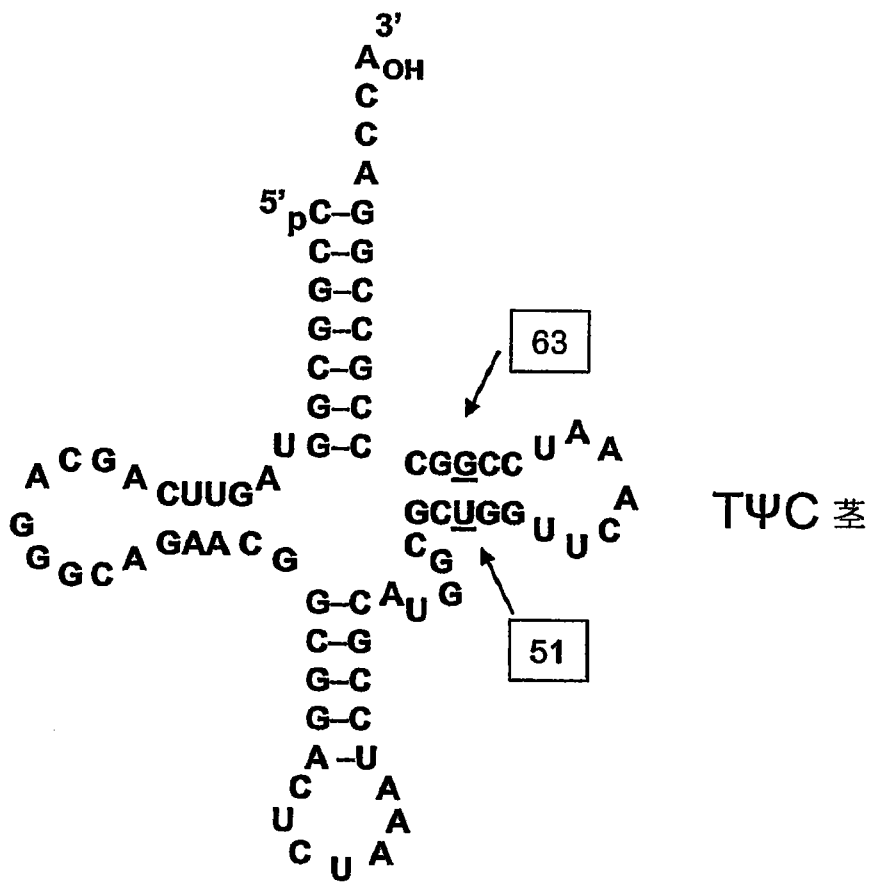


图1

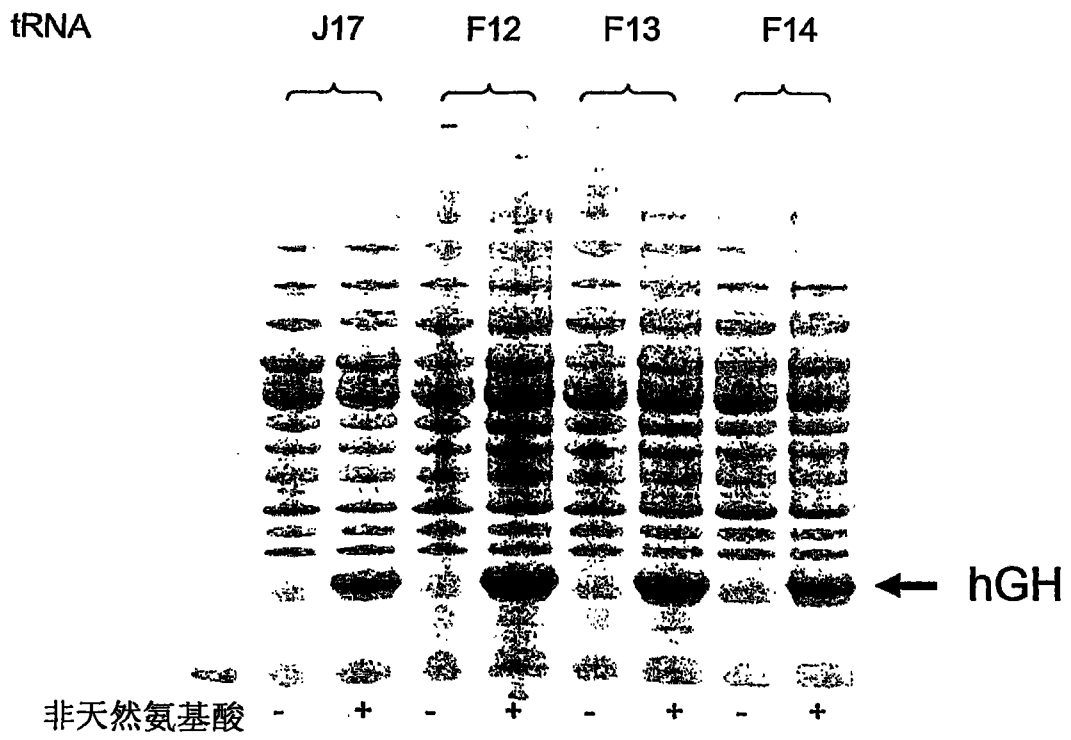


图2

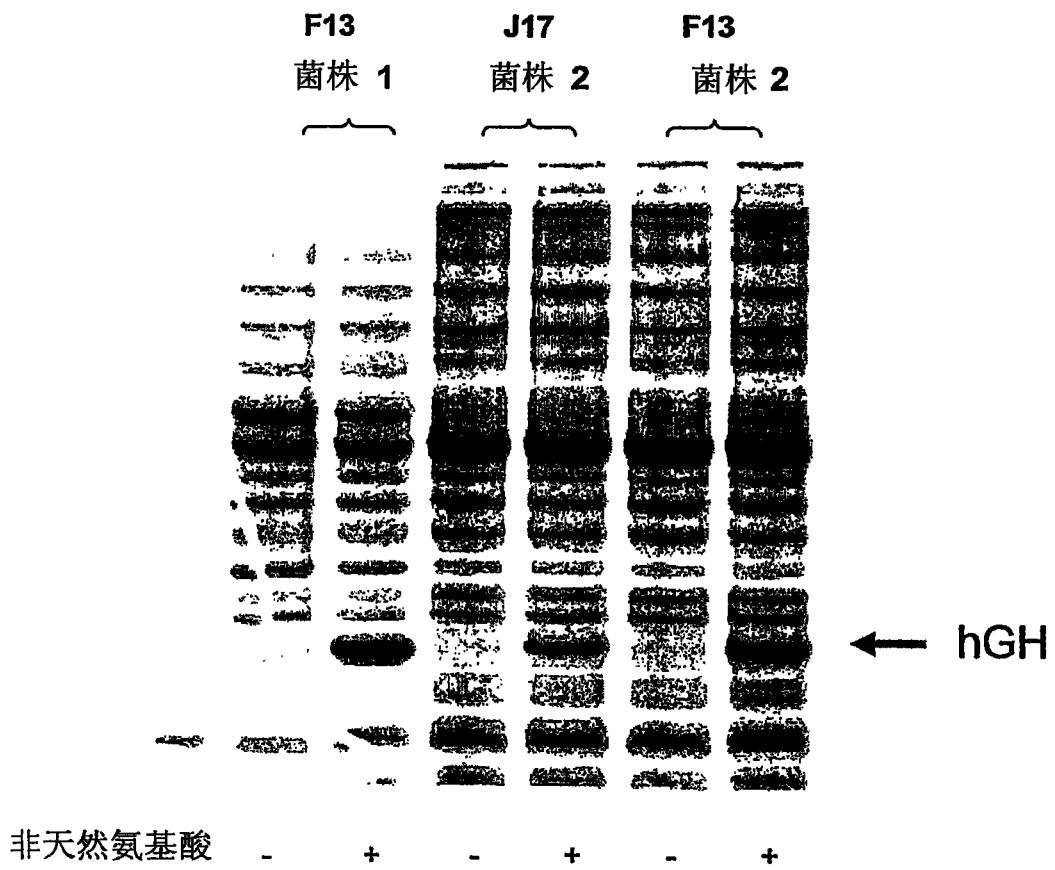


图3